POLITECNICO
DI MILANO
Master in Mathematical
Engineering

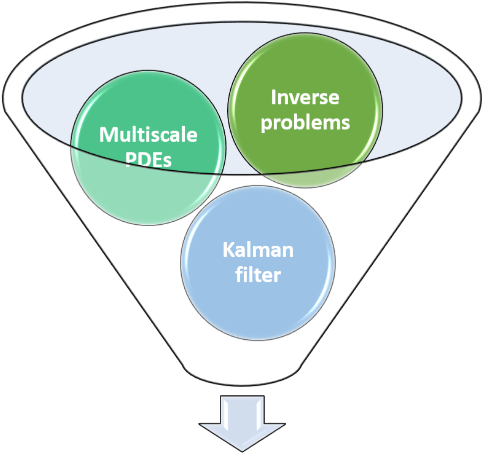
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Master in Computational
Science and Engineering

Ensemble Kalman filter for multiscale inverse problems

Assyr Abdulle, Giacomo Garegnani, Sandro Salsa

MATHICSE Retreat
11–13 June 2019

Main ingredients



ENSEMBLE KALMAN FILTER FOR MULTISCALE INVERSE PROBLEMS

Outline

1. Inverse problems
2. Ensemble Kalman method
3. Multiscale inverse problems
4. Numerical experiments

Inverse problem

$$\text{Find} \quad u \in X \quad \text{given} \quad y = \mathcal{G}(u) + \quad \eta \quad \in Y$$

Inverse problem

$$\text{Find } \underbrace{u \in X}_{\text{unknown}} \text{ given } \underbrace{y = \mathcal{G}(u) + \overbrace{\eta}^{\text{noise}}}_{\text{observation}} \in Y$$

- X, Y Hilbert spaces
- $\eta \sim \mathcal{N}(0, \Gamma)$

Inverse problem

$$\text{Find } \underbrace{u \in X}_{\text{unknown}} \text{ given } \underbrace{y = \mathcal{G}(u) + \overbrace{\eta}^{\text{noise}}}_{\text{observation}} \in Y$$

- X, Y Hilbert spaces
- $\eta \sim \mathcal{N}(0, \Gamma)$

- $\mathcal{G}\colon X \to Y$ **forward operator**, $\mathcal{G} = \mathcal{O} \circ \mathcal{S}$
  - $\mathcal{O}\colon H^1(\Omega) \to Y$ observation operator
  - $\mathcal{S}\colon X \to H^1(\Omega)$ solution operator of PDE

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

Overview of different techniques

Given a prior μ_0, consider two families of techniques

- Tikhonov methods $\longrightarrow$ solution is a **point**

$$u^* = \operatorname*{argmin}_{u \in X} \Big(\underbrace{\Psi(u; y)}_{\text{misfit functional}} + \underbrace{\mathcal{R}(u; \mu_0)}_{\text{regularization term}} \Big)$$

Example

Least squares with Tikhonov–Phillips regularization ($\mu_0 = \mathcal{N}(\bar{u}, C)$)

$$u^* = u_{\mathrm{TP}} = \operatorname*{argmin}_{u \in X} \left(\|y - \mathcal{G}(u)\|_\Gamma^2 + \|u - \bar{u}\|_C^2 \right)$$

Overview of different techniques

Given a prior μ_0, consider two families of techniques

- Bayesian methods $\longrightarrow$ solution is a **distribution**

$$\frac{d\mu^y}{d\mu_0}(u) \propto \exp\Big(-\underbrace{\Phi(u;y)}_{\text{potential}}\Big)$$

Example

Gaussian noise and prior measure ($\eta \sim \mathcal{N}(0,\Gamma)$, $\mu_0 = \mathcal{N}(\bar{u}, C)$)

$$\mu^y(u) \propto \exp\left\{-\frac{1}{2}\left(\|y-\mathcal{G}(u)\|_\Gamma^2 + \|u-\bar{u}\|_C^2\right)\right\}$$

Outline

Kalman filter

Kalman filter estimates state $z_n \in Z$ of a linear dynamical system

$$z_{n+1} = G z_n$$

given observations $y_n \in Y$

$$y_n = H z_n + \eta_n$$

where

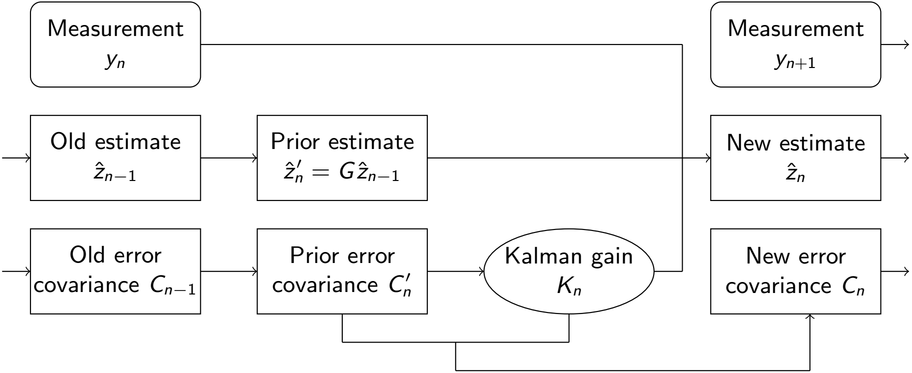
- G transition matrix
- H observation matrix
- $\eta_n \sim \mathcal{N}(0, \Gamma)$ noise

Kalman filter

$$z_{n+1} = Gz_n$$
$$y_n = Hz_n + \eta_n$$

Compute the estimation $\hat{z}_n$ in order to minimize

$$\mathrm{MSE} = \mathbb{E}[(z_n - \hat{z}_n)^T(z_n - \hat{z}_n)]$$



Kalman filter for inverse problems

Idea (see Iglesias et al. (2013))

Apply Kalman filter to solve inverse problems

Kalman filter for inverse problems

Idea (see Iglesias et al. (2013))

Apply Kalman filter to solve inverse problems

Issue

There are no dynamics

Kalman filter for inverse problems

Idea (see Iglesias et al. (2013))

Apply Kalman filter to solve inverse problems

Issue

There are no dynamics

Introduce artificial dynamics

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

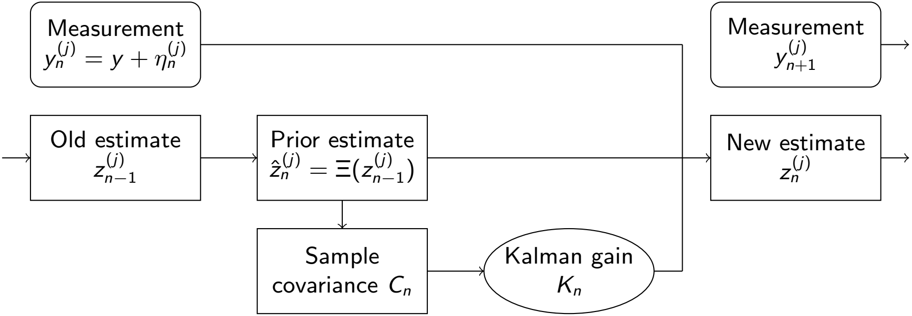
where

- $z_n = \begin{bmatrix} u_n & v_n \end{bmatrix}^T \in Z = X \times Y$
- $\Xi(z_n) = \begin{bmatrix} u_n & \mathcal{G}(u_n) \end{bmatrix}^T$, $H = \begin{bmatrix} 0 & I \end{bmatrix}^T$

Observations are obtained by randomizing y: $y_n = y + \eta_n$

Ensemble Kalman filter

- Initial ensemble z_0 : $z_0^{(j)} = \begin{bmatrix} \psi^{(j)} & \mathcal{G}(\psi^{(j)}) \end{bmatrix}^T$, $\psi^{(j)} \sim \mu_0$, $j = 1, \dots, J$
- Iterate for $n = 1, \dots, N$



- Compute the solution $u_{\mathrm{EnKF}} = \frac{1}{J} \sum_{j=1}^{J} u_N^{(j)}$

Ensemble Kalman filter

Properties

- $u_{\mathrm{EnKF}} \in \mathrm{span}\,\left(\{\psi^{(j)}\}_{j=1}^{J}\right)$
- If $\mathcal{G}$ **linear** then $u_{\mathrm{EnKF}}(N = 1) \to u_{\mathrm{TP}}$ as $J \to \infty$
- Easily parallelizable

Ensemble Kalman filter

Properties

- $u_{\mathrm{EnKF}} \in \mathrm{span}\left(\{\psi^{(j)}\}_{j=1}^{J}\right)$
- If $\mathcal{G}$ **linear** then $u_{\mathrm{EnKF}}(N=1) \to u_{\mathrm{TP}}$ as $J \to \infty$
- Easily parallelizable

Warning

$J \cdot N$ evaluations of the forward operator $\mathcal{G}$ required

Outline

Multiscale inverse problems

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

where

$$\mathcal{G}^\varepsilon = \mathcal{O} \circ \mathcal{S}^\varepsilon$$

and $\mathcal{S}^\varepsilon$ is the solution operator of a multiscale PDE

$$\begin{cases} -\nabla \cdot (A_u^\varepsilon \nabla p^\varepsilon) = f & \text{in } \Omega \\ p^\varepsilon = g & \text{on } \partial\Omega \end{cases} \qquad A_u^\varepsilon(x) = A\left(u(x), \frac{x}{\varepsilon}\right)$$

Multiscale inverse problems

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

where

$$\mathcal{G}^\varepsilon = \mathcal{O} \circ \mathcal{S}^\varepsilon$$

and $\mathcal{S}^\varepsilon$ is the solution operator of a multiscale PDE

$$\begin{cases} -\nabla \cdot (A_u^\varepsilon \nabla p^\varepsilon) = f & \text{in } \Omega \\ p^\varepsilon = g & \text{on } \partial\Omega \end{cases} \qquad A_u^\varepsilon(x) = A\left(u(x), \frac{x}{\varepsilon}\right)$$

Issue

Evaluation of $\mathcal{G}^\varepsilon$ computationally very expensive if $\varepsilon \ll 1$

Homogenization

Theorem

There exist A_u^0 and p^0 such that

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

where p^0 is the solution of

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

Homogenization

Theorem

There exist A_u^0 and p^0 such that

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

where p^0 is the solution of

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

Idea (see Nolen et al. (2012))

Replace $\mathcal{G}^\varepsilon$ with $\mathcal{G}_h^0 = \mathcal{O} \circ \mathcal{S}_h^0$, where $\mathcal{S}_h^0$ is the solution operator corresponding to the FEM discretizazion of the homogenized problem

A one-dimensional example

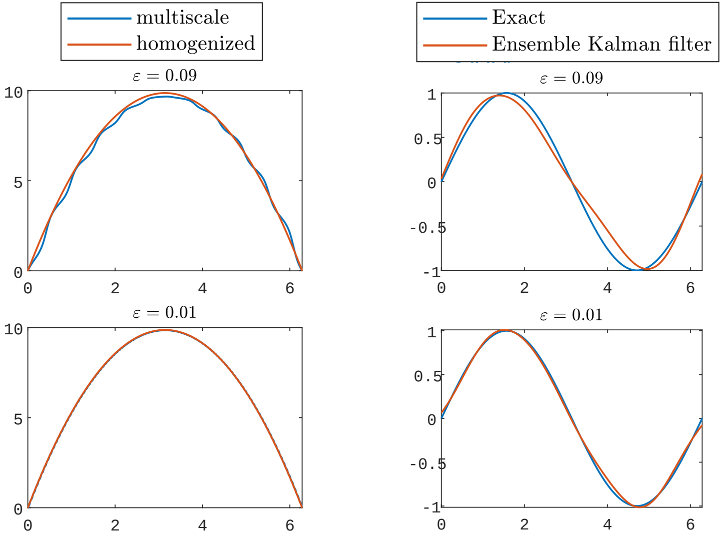


Figure : Left: solution of the PDE. Right: unknown of the inverse problem.

Convergence analysis

- $u_N^\varepsilon = \{u_N^{\varepsilon(j)}\}_{j=1}^J$ ensemble generated by EnKF with $\mathcal{G}^\varepsilon$
- $u_{N,h}^0 = \{u_{N,h}^{0(j)}\}_{j=1}^J$ ensemble generated by EnKF with $\mathcal{G}_h^0$
- norm of the ensemble defined by

$$\|u\| := \frac{1}{J}\sum_{j=1}^{J}\|u^{(j)}\|_X$$

Convergence analysis

- $u_N^\varepsilon = \{u_N^{\varepsilon(j)}\}_{j=1}^J$ ensemble generated by EnKF with $\mathcal{G}^\varepsilon$
- $u_{N,h}^0 = \{u_{N,h}^{0(j)}\}_{j=1}^J$ ensemble generated by EnKF with $\mathcal{G}_h^0$
- norm of the ensemble defined by

$$\|u\| := \frac{1}{J}\sum_{j=1}^{J}\|u^{(j)}\|_X$$

Main assumptions

- $\mathcal{O}$ Lipschitz from L^2 to Y
- A elliptic and Lipschitz from X to L^∞
- Algorithm stable

Convergence analysis

Theorem

Under the previous assumptions

$$\mathbb{E}\left[\|u_N^{\varepsilon} - u_{N,h}^{0}\|\right] \leq C(\varepsilon + h^{s+1})$$

where $s = \min\{r, q\}$, $p^0 \in H^{q+1}(\Omega)$ *and* r *degree of the FEM basis*

Convergence analysis

Theorem

Under the previous assumptions

$$\mathbb{E}\left[\|u_N^{\varepsilon} - u_{N,h}^0\|\right] \leq C(\varepsilon + h^{s+1})$$

where $s = \min\{r, q\}$, $p^0 \in H^{q+1}(\Omega)$ *and* r *degree of the FEM basis*

Idea of the proof

- Let $u_N^0 = \{u_N^{0(j)}\}_{j=1}^J$ be the ensemble generated by EnKF with $\mathcal{G}^0$
- By triangle inequality

$$\mathbb{E}\left[\|u_N^{\varepsilon} - u_{N,h}^0\|\right] \leq \underbrace{\mathbb{E}\left[\|u_N^{\varepsilon} - u_N^0\|\right]}_{\substack{\text{multiscale} \\ \text{convergence} \\ \varepsilon \to 0 \\ \Downarrow \\ \leq C_1 \varepsilon}} + \underbrace{\mathbb{E}\left[\|u_N^0 - u_{N,h}^0\|\right]}_{\substack{\text{FEM} \\ \text{convergence} \\ h \to 0 \\ \Downarrow \\ \leq C_2 h^{s+1}}}$$

Bayesian interpretation of the ensemble Kalman filter

Posterior distribution approximated by sum of Dirac masses (see Schillings and Stuart (2017))

$$\mu_N = \frac{1}{J} \sum_{j=1}^{J} \delta_{u_N^{(j)}}$$

- $\mu_N^\varepsilon = \frac{1}{J} \sum_{j=1}^{J} \delta_{u_N^{\varepsilon(j)}}$ posterior generated by EnKF with $\mathcal{G}^\varepsilon$
- $\mu_{N,h}^0 = \frac{1}{J} \sum_{j=1}^{J} \delta_{u_{N,h}^{0(j)}}$ posterior generated by EnKF with $\mathcal{G}_h^0$

Bayesian interpretation of the ensemble Kalman filter

Posterior distribution approximated by sum of Dirac masses (see Schillings and Stuart (2017))

$$\mu_N = \frac{1}{J}\sum_{j=1}^{J}\delta_{u_N^{(j)}}$$

- $\mu_N^{\varepsilon} = \frac{1}{J}\sum_{j=1}^{J}\delta_{u_N^{\varepsilon(j)}}$ posterior generated by EnKF with $\mathcal{G}^{\varepsilon}$
- $\mu_{N,h}^{0} = \frac{1}{J}\sum_{j=1}^{J}\delta_{u_{N,h}^{0(j)}}$ posterior generated by EnKF with $\mathcal{G}_h^{0}$

Warning

The posterior distributions are random probability measures

Convergence of the posterior distributions

Theorem

Under the same assumptions of the convergence result

$$\mu_N^\varepsilon - \mu_{N,h}^0 \xrightarrow{L^1} 0 \qquad as \qquad \varepsilon, h \to 0$$

which means

$$\mathbb{E}\left[\left|\int f \,\mathrm{d}\mu_N^\varepsilon - \int f \,\mathrm{d}\mu_{N,h}^0\right|\right] \to 0$$

for all bounded and continuous functions $f \in C_B^0$

Convergence of the posterior distributions

Theorem

Under the same assumptions of the convergence result

$$\mu_N^{\varepsilon} - \mu_{N,h}^{0} \xrightarrow{L^1} 0 \qquad \text{as} \qquad \varepsilon, h \to 0$$

which means

$$\mathbb{E}\left[\left|\int f \, \mathrm{d}\mu_N^{\varepsilon} - \int f \, \mathrm{d}\mu_{N,h}^{0}\right|\right] \to 0$$

for all bounded and continuous functions $f \in C_B^0$

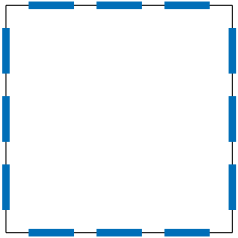
Idea of the proof

- Exploit properties of Wasserstein distance to quantify distance between discrete measures
- Apply convergence result presented before

Outline

Observations and exact solution

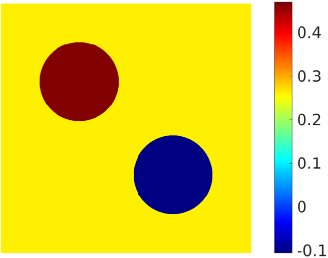
Setting from Abdulle and Di Blasio (2018)



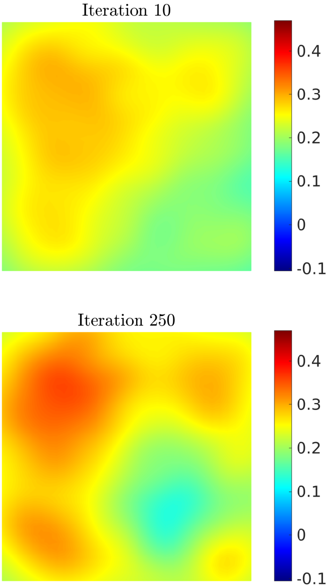
Measurements ($i = 1, \ldots, 12$)

$$y_i = \int_{\Gamma_i} A^{\varepsilon} \nabla p_k^{\varepsilon} \cdot \nu \phi_i ds + \eta_i$$

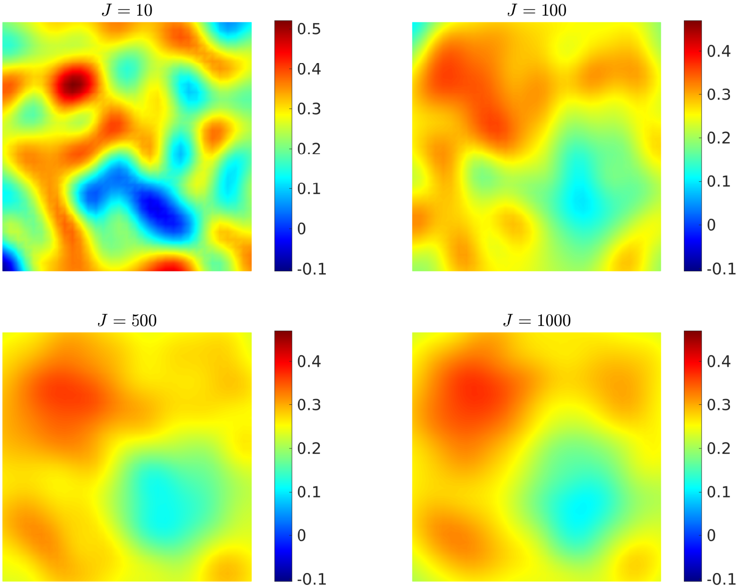
Exact unknown



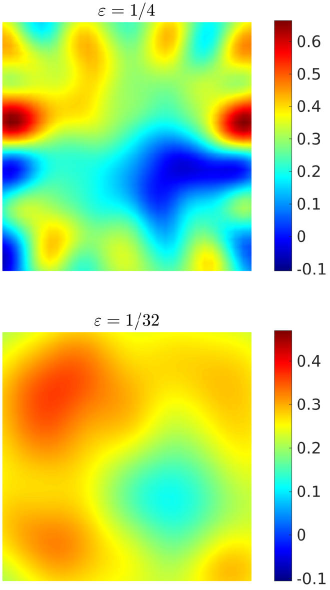
Sensitivity w.r.t. N



Sensitivity w.r.t. J



Sensitivity w.r.t. ε



Modelling error

Idea (see Kaipio and Somersalo (2005))

The observation can be seen as data generated by the discrete homogenized operator affected by two sources of error

$$y = \mathcal{G}_h^0(u^*) + \underbrace{[\mathcal{G}^\varepsilon(u^*) - \mathcal{G}_h^0(u^*)]}_{=\ \mathcal{E}} + \eta$$

Modelling error

Idea (see Kaipio and Somersalo (2005))

The observation can be seen as data generated by the discrete homogenized operator affected by two sources of error

$$y = \mathcal{G}_h^0(u^*) + \underbrace{[\mathcal{G}^\varepsilon(u^*) - \mathcal{G}_h^0(u^*)]}_{=\ \mathcal{E}} + \eta$$

Assume that

$$\mathcal{E} \sim \mathcal{N}(m, \Sigma)$$

then $\mathcal{E} = m + \zeta$ with $\zeta \sim \mathcal{N}(0, \Sigma)$ and

$$\underbrace{y - m}_{=\ \tilde{y}} = \mathcal{G}_h^0(u^*) + \underbrace{\zeta + \eta}_{=\ \tilde{\eta}}$$

Estimation of the sample mean m and covariance Σ

Two approaches can be followed

1. **offline:**
    - sample $\{u_i\}_{i=1}^{N_\mathcal{E}}$ from the prior distribution μ_0
    - compute $\mathcal{E}_i = \mathcal{G}^\varepsilon(u_i) - \mathcal{G}_h^0(u_i)$
    - $m = \frac{1}{N_\mathcal{E}} \sum_{i=1}^{N_\mathcal{E}} \mathcal{E}_i$ and $\Sigma = \frac{1}{N_\mathcal{E}} \sum_{i=1}^{N_\mathcal{E}} (\mathcal{E}_i - m)(\mathcal{E}_i - m)^T$

Estimation of the sample mean m and covariance Σ

Two approaches can be followed

1. **offline:**
   - sample $\{u_i\}_{i=1}^{N_\mathcal{E}}$ from the prior distribution μ_0
   - compute $\mathcal{E}_i = \mathcal{G}^\varepsilon(u_i) - \mathcal{G}_h^0(u_i)$
   - $m = \frac{1}{N_\mathcal{E}} \sum_{i=1}^{N_\mathcal{E}} \mathcal{E}_i$ and $\Sigma = \frac{1}{N_\mathcal{E}} \sum_{i=1}^{N_\mathcal{E}} (\mathcal{E}_i - m)(\mathcal{E}_i - m)^T$
2. **online:** subdivide the EnKF in $\mathcal{L}$ levels, each one with N^ℓ iterations
   - sample $\{u_i^\ell\}_{i=1}^{N_\mathcal{E}^\ell}$ from the distribution μ_0^ℓ $(\mu_0^{\ell+1} = \mu_{N^\ell}^\ell)$
   - compute $\mathcal{E}_i^\ell = \mathcal{G}^\varepsilon(u_i^\ell) - \mathcal{G}_h^0(u_i^\ell)$
   - $m^\ell = \frac{1}{N_\mathcal{E}^\ell} \sum_{i=1}^{N_\mathcal{E}^\ell} \mathcal{E}_i^\ell$ and $\Sigma^\ell = \frac{1}{N_\mathcal{E}^\ell} \sum_{i=1}^{N_\mathcal{E}^\ell} (\mathcal{E}_i^\ell - m^\ell)(\mathcal{E}_i^\ell - m^\ell)^T$

Estimation of the sample mean m and covariance Σ

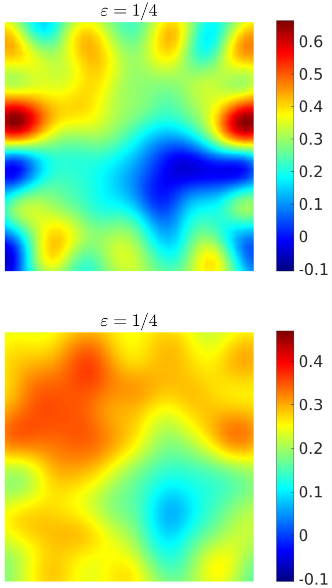
Two approaches can be followed

1. **offline:**
   - sample $\{u_i\}_{i=1}^{N_\varepsilon}$ from the prior distribution μ_0
   - compute $\mathcal{E}_i = \mathcal{G}^\varepsilon(u_i) - \mathcal{G}_h^0(u_i)$
   - $m = \frac{1}{N_\varepsilon}\sum_{i=1}^{N_\varepsilon} \mathcal{E}_i$ and $\Sigma = \frac{1}{N_\varepsilon}\sum_{i=1}^{N_\varepsilon}(\mathcal{E}_i - m)(\mathcal{E}_i - m)^T$
2. **online:** subdivide the EnKF in $\mathcal{L}$ levels, each one with N^ℓ iterations
   - sample $\{u_i^\ell\}_{i=1}^{N_\varepsilon^\ell}$ from the distribution μ_0^ℓ $(\mu_0^{\ell+1} = \mu_{N^\ell}^\ell)$
   - compute $\mathcal{E}_i^\ell = \mathcal{G}^\varepsilon(u_i^\ell) - \mathcal{G}_h^0(u_i^\ell)$
   - $m^\ell = \frac{1}{N_\varepsilon^\ell}\sum_{i=1}^{N_\varepsilon^\ell} \mathcal{E}_i^\ell$ and $\Sigma^\ell = \frac{1}{N_\varepsilon^\ell}\sum_{i=1}^{N_\varepsilon^\ell}(\mathcal{E}_i^\ell - m^\ell)(\mathcal{E}_i^\ell - m^\ell)^T$

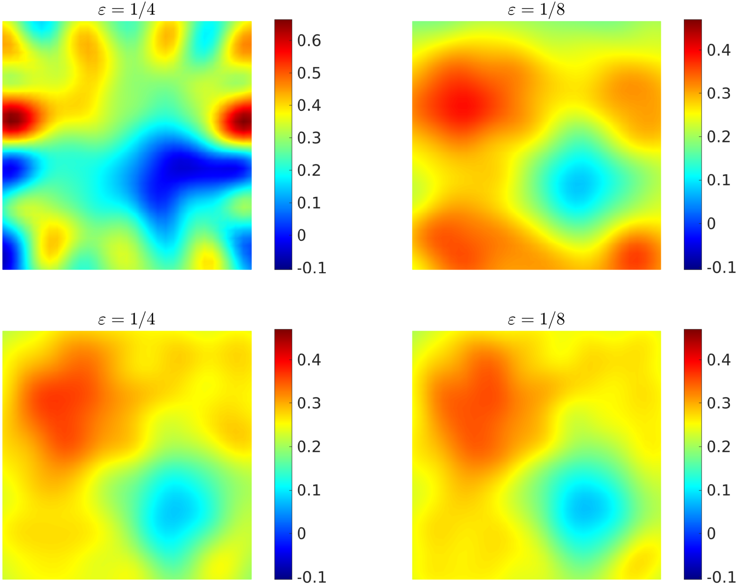
Warning

Online approximation is better, but computationally more expensive

Offline modelling error estimation



Online modelling error estimation



Conclusions

- The ensemble Kalman filter can be used for multiscale inverse problems
- The multiscale forward operator can be replaced by the homogenized one, taking into account the modelling error when ε is big
- The ensemble Kalman filter is easily parallelizable to reduce the computational time

References

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

Iglesias, M. A., Law, K. J. H., and Stuart, A. M. (2013). Ensemble Kalman methods for inverse problems. *Inverse Problems*, 29(4):045001, 20.

Kaipio, J. and Somersalo, E. (2005). *Statistical and computational inverse problems*, volume 160 of *Applied Mathematical Sciences*. Springer-Verlag, New York.

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.

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