Covariance-Informed Subspace: a Gradient-free Dimension Reduction for Adaptive Bayesian Inference

MASCOTNUM2025 - PhD Day

PhD student: Nadège Polette[1,2],
Supervisors:
Alexandrine Gesret[1], Pierre Sochala[2], Olivier Le Maître[3]

[1]Mines Paris PSL, Geosciences center, Fontainebleau, France
[2]CEA, DAM, DIF, F-91297 Arpajon, France
[3]CNRS, CMAP, IPP École Polytechnique, Palaiseau, France

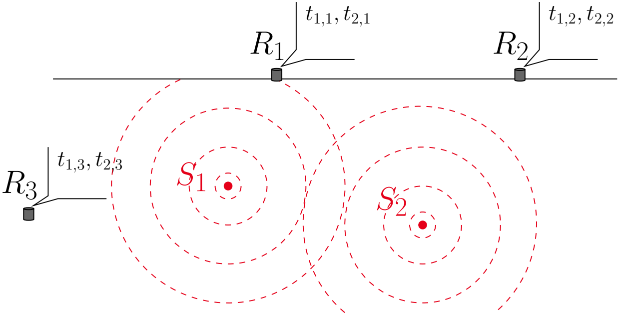
Table of contents

1. Context: Detection and analysis of seismic events

Global scale
- International treaties (CTBT, NTP)
- Environment monitoring (IMS)

Regional scale
- Tsunami and earthquake alerts
- Risk prevention

Local scale
- Subsurface knowledge
- Exploitation

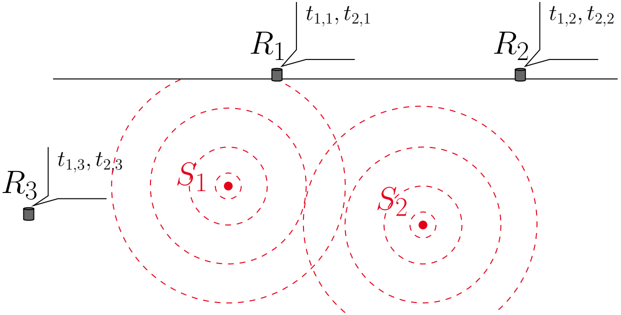


1. Context: Detection and analysis of seismic events

Global scale
- International treaties (CTBT, NTP)
- Environment monitoring (IMS)

Regional scale
- Tsunami and earthquake alerts
- Risk prevention

Local scale
- Subsurface knowledge
- Exploitation



$$F(S) = \boldsymbol{d}$$

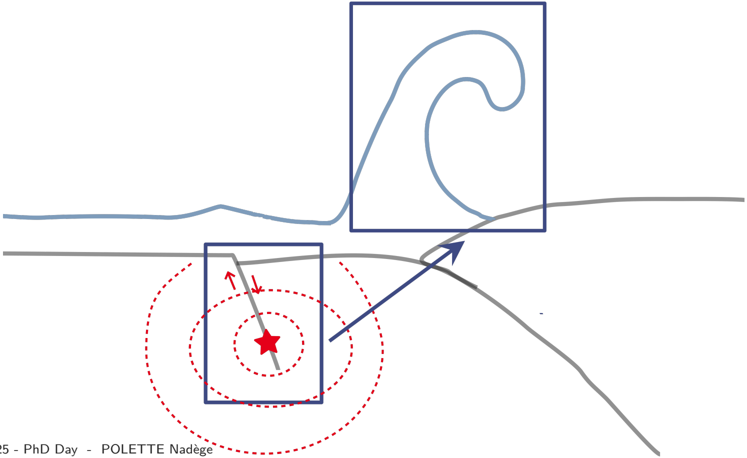
F: forward model
S: source parameters
$\boldsymbol{d}$: data

Objective - retrieve S from $\boldsymbol{d}$

- fast
- accurately
- with uncertainties

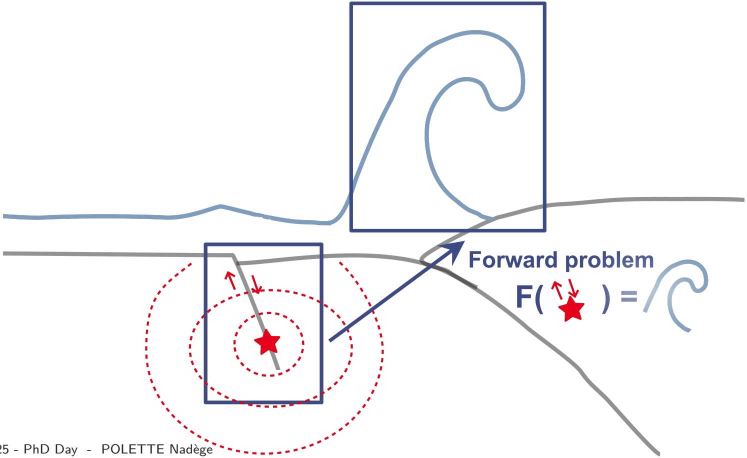
1. Context: Uncertainty on physical parameters

- Earthquakes could cause tsunamis within minutes/hours



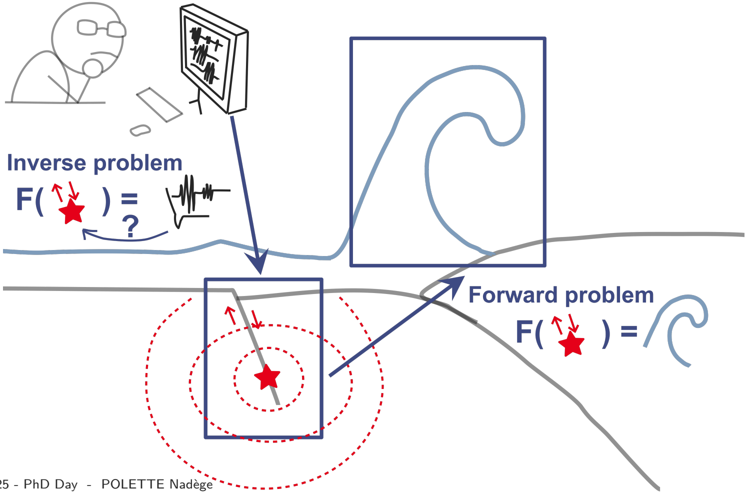
1. Context: Uncertainty on physical parameters

- Earthquakes could cause tsunamis within minutes/hours
- Source parameters → Simulation



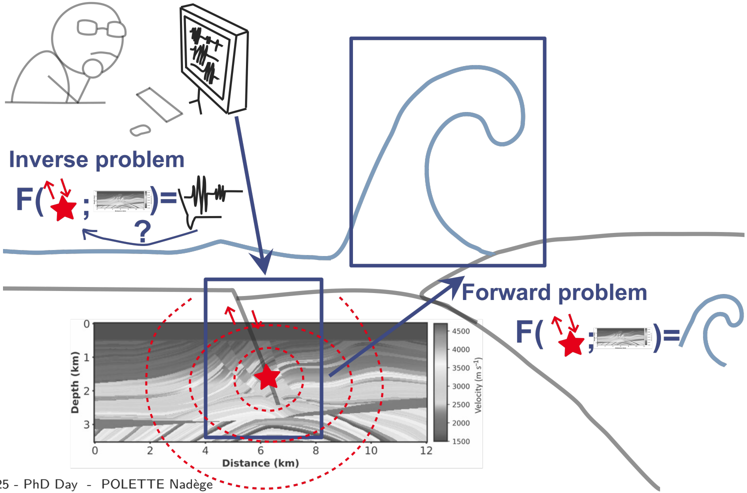
1. Context: Uncertainty on physical parameters

- Earthquakes could cause tsunamis within minutes/hours
- Data → Source parameters → Simulation

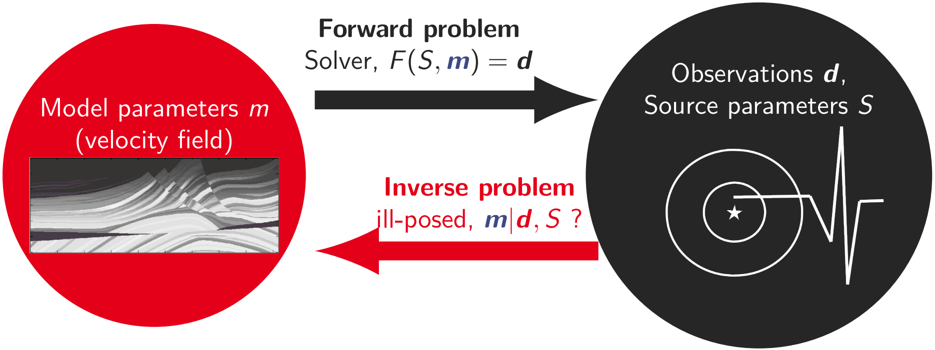


1. Context: Uncertainty on physical parameters

- Earthquakes could cause tsunamis within minutes/hours
- Data $\rightarrow$ Source parameters $\rightarrow$ Simulation
- Poorly known parameters, e.g. the velocity field



1. Context: Inverse problem



Objective: **to characterize the velocity field** m **and its uncertainty** from indirect observations $\boldsymbol{d}$
$\Rightarrow$ to find the probability distribution of the field knowing the observations $\pi_{\text{post}}(m|\boldsymbol{d}^{\text{obs}})$

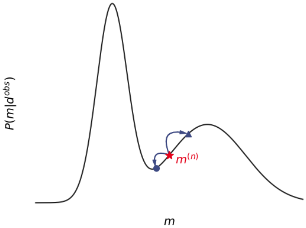
Tarantola, *SIAM*, 2005; **Noble et al.**, *GJI*, 2014

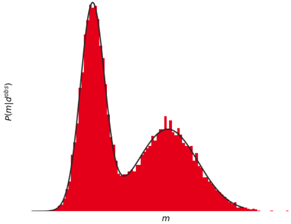
1. Context: Bayesian inference and Markov chain Monte Carlo

Proposition $m^{(i)}$ → **Forward model** Solver $F(m; S) = \boldsymbol{d}$ → Comparison $\boldsymbol{d}$ versus $\boldsymbol{d}^{\text{obs}}$ (data)

Bayes' rule: $\pi_{\text{post}}(m|\boldsymbol{d}^{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}^{\text{obs}}|m)\, \pi_{\text{prior}}(m)$

Accept/Reject criterion → New iteration $i \leftarrow i+1$ → Proposition $m^{(i)}$





Sivia and Skilling, *Oxford*, 2006; **Doucet et al.**, *Springer NY*, 2013

1. Context: Bayesian inference and Markov chain Monte Carlo

Proposition $m^{(i)}$

m infinite dimensional
$\rightarrow$ dimension reduction

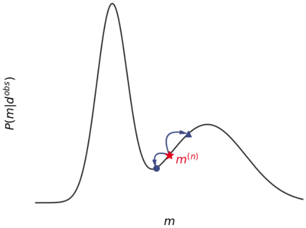
Forward model
Solver $F(m;S) = \boldsymbol{d}$

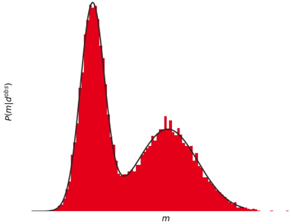
Comparison
$\boldsymbol{d}$ versus $\boldsymbol{d}^{\text{obs}}$ (data)

Bayes' rule: $\pi_{\text{post}}(m|\boldsymbol{d}^{\text{obs}}) \propto \overbrace{\mathcal{L}(\boldsymbol{d}^{\text{obs}}|m)}\,\pi_{\text{prior}}(m)$

Accept/Reject criterion

New iteration
$i \leftarrow i+1$





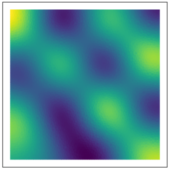
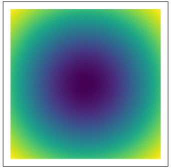
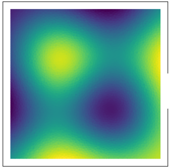
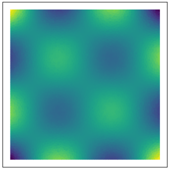
Sivia and Skilling, *Oxford*, 2006; **Doucet et al.**, *Springer NY*, 2013

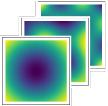
Table of contents

2. Field parametrization: Modal representation

Assumption - $m \sim \mathcal{N}(0, K)$

Karhunen–Loève decomposition - $m(\boldsymbol{x}) = \sum_{i=1}^{n} \sqrt{\lambda_i} u_i(\boldsymbol{x}) \eta_i, \ \boldsymbol{x} \in \Omega$

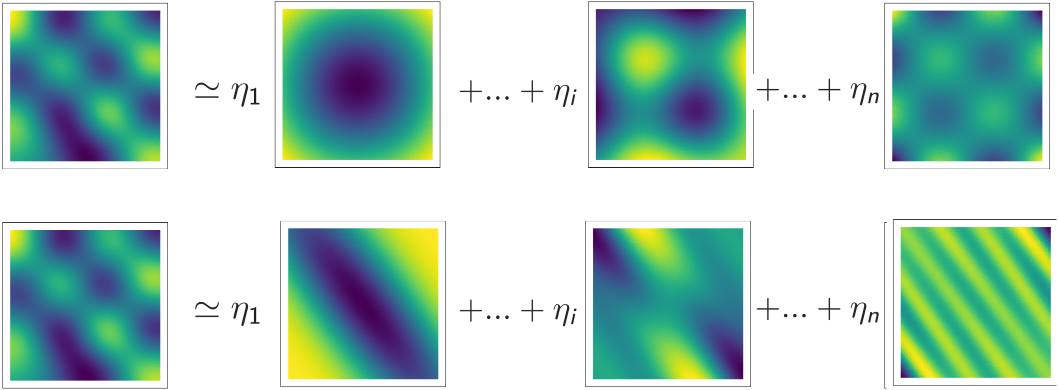
 $\simeq \eta_1$  $+ \ldots + \eta_i$  $+ \ldots + \eta_n$ 

-  obtained using eigendecomposition of K: $\int_\Omega K(\boldsymbol{x}, \boldsymbol{y}) u_i(\boldsymbol{x}) d\boldsymbol{x} = \lambda_i u_i(\boldsymbol{y})$
- $\pi_{\text{post}}(m|\boldsymbol{d}^{\text{obs}}) \Rightarrow \pi_{\text{post}}(\boldsymbol{\eta}|\boldsymbol{d}^{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\eta})\pi(\boldsymbol{\eta}), \quad \boldsymbol{\eta} \sim \mathcal{N}(0, \mathrm{I}_n)$

Marzouk and Najm, *JCP*, 2009

2. Field parametrization: Kernel choice

Problem - the choice of K is not always trivial



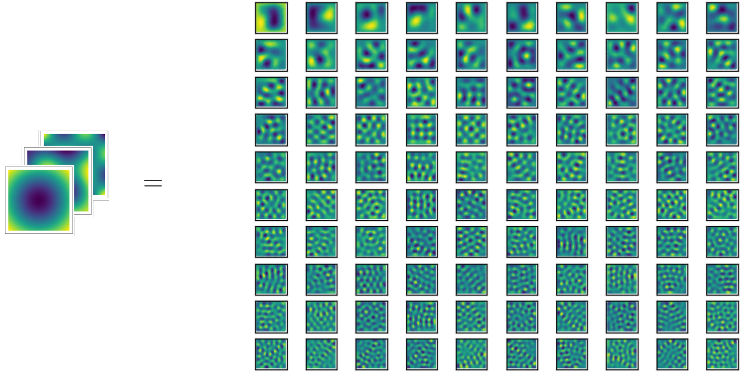
- $\pi_{\text{post}}(m|\boldsymbol{d}^{\text{obs}}) \Rightarrow \pi_{\text{post}}(\boldsymbol{\eta},\boldsymbol{q}|\boldsymbol{d}^{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\eta},\boldsymbol{q})\pi(\boldsymbol{\eta})\pi(\boldsymbol{q}), \quad \boldsymbol{\eta}\sim\mathcal{N}(0,\mathrm{I}_n),\ \boldsymbol{q}\sim\pi_q$

Polette, Le Maître, Sochala, Gesret, *JCP*, 2025

Table of contents

3. Dimension reduction: Context

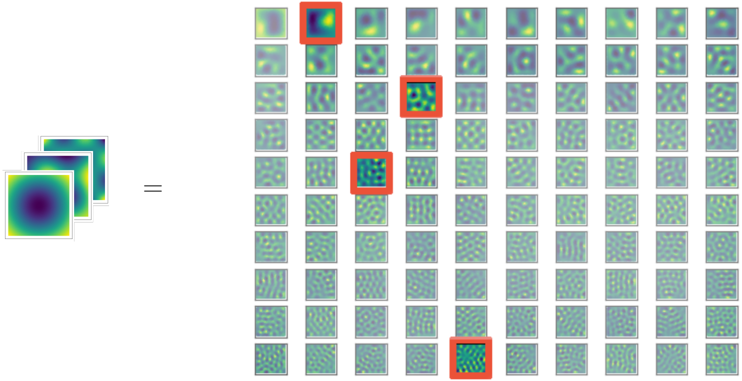
Problem - Parametrisation may require a large number of modes



$\Rightarrow$ Computationally expensive

3. Dimension reduction: Context

Idea - Select only a sub-collection

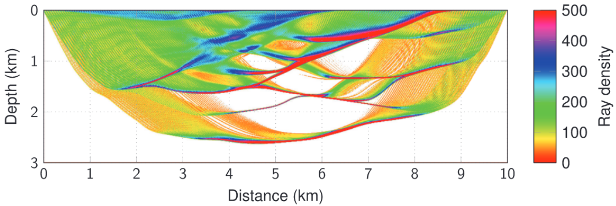


$\Rightarrow$ How to choose the selected modes ?

3. Dimension reduction: Informed subspace

Assumption - the data are informative only on a low-dimensional subspace

$$\pi_{\text{post}}(\boldsymbol{\eta}|\boldsymbol{d}^{\text{obs}}) \simeq \widetilde{\pi}_{\text{post}}(\boldsymbol{\eta}|\boldsymbol{d}^{\text{obs}}) \propto \widetilde{\mathcal{L}}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\eta}_r)\pi(\boldsymbol{\eta}_r)\pi(\boldsymbol{\eta}_\perp|\boldsymbol{\eta}_r), \quad \boldsymbol{\eta} = A_r\boldsymbol{\eta}_r + A_\perp\boldsymbol{\eta}_\perp$$



Data limitations: ray density map (**Luu et al.**, *Geoph. Prosp.*, 2020)

Constantine et al., *SIAM JSC*, 2016; **Cui et al.**, *Inv. Prob.*, 2014; **Zahm et al.**, *Math. Comp.*, 2022;

3. Dimension reduction: Informed subspace

Proposition $\eta_r^{(i)}$

Approximated likelihood
$\widetilde{\mathcal{L}}(\boldsymbol{d}^{\mathrm{obs}}|\eta_r)$

Comparison
$\boldsymbol{d}$ versus $\boldsymbol{d}^{\mathrm{obs}}$ (data)

Bayes' rule: $\widetilde{\pi}_{\mathrm{post}}(\eta_r|\boldsymbol{d}^{\mathrm{obs}}) \propto \widetilde{\mathcal{L}}(\boldsymbol{d}^{\mathrm{obs}}|\eta_r)\,\pi_{\mathrm{prior}}(\eta_r)$

Accept/Reject criterion

New iteration
$i \leftarrow i+1$

- $\eta_\perp$ is sampled according to $\pi_{\mathrm{prior}}(\eta_\perp|\eta_r)$
- **Approximated likelihood**

$$\widetilde{\mathcal{L}}(\boldsymbol{d}^{\mathrm{obs}}|\eta_r) = \mathbb{E}_{\pi_{\mathrm{prior}}}(\mathcal{L}|\eta_r) = \int_{\Xi_\perp} \mathcal{L}(\boldsymbol{d}^{\mathrm{obs}}|\boldsymbol{\eta})\pi_{\mathrm{prior}}(\eta_\perp|\eta_r)d\eta_\perp$$

  - optimal with respect to (i) the $L^2\pi$-norm, (ii) the Kullback–Leibler divergence [Zahm2022, Section 2], and more generally (iii) all expected Bregman divergences [Banerjee2005]

Zahm et al., *Math. Comp.*, 2022; **Banerjee et al.**, *IEE Trans. Inform. Theory 51*, 2005

3. Dimension reduction: Informed subspace

Assumption - the data are informative only on a low-dimensional subspace

$$\pi_{\text{post}}(\boldsymbol{\eta}|\boldsymbol{d}^{\text{obs}}) \simeq \widetilde{\pi}_{\text{post}}(\boldsymbol{\eta}|\boldsymbol{d}^{\text{obs}}) \propto \widetilde{\mathcal{L}}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\eta}_r)\pi(\boldsymbol{\eta}_r)\pi(\boldsymbol{\eta}_\perp|\boldsymbol{\eta}_r), \quad \boldsymbol{\eta} = A_r\boldsymbol{\eta}_r + A_\perp\boldsymbol{\eta}_\perp$$

$\Rightarrow$ State-of-the-art methods are *gradient-based*.

- $H = \int_\Xi \nabla \log \mathcal{L}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\eta}) \left(\nabla \log \mathcal{L}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\eta})\right)^\top \nu(\boldsymbol{\eta})d\boldsymbol{\eta}$
- The reduction relies on the dominant eigenspace of the pencil $(H, C_{\text{prior}}^{-1} = \mathrm{I}_n)$:

$$[u_1 \dots u_r], \quad \lambda_1 \geqslant \dots \geqslant \lambda_n, \quad Hu_i = \lambda_i C_{\text{prior}}^{-1} u_i$$

Constantine et al., *SIAM JSC*, 2016; **Cui et al.**, *Inv. Prob.*, 2014; **Zahm et al.**, *Math. Comp.*, 2022;

3. Dimension reduction: Informed subspace

Assumption - the data are informative only on a low-dimensional subspace

$$\pi_{\text{post}}(\boldsymbol{\eta}|\boldsymbol{d}^{\text{obs}}) \simeq \widetilde{\pi}_{\text{post}}(\boldsymbol{\eta}|\boldsymbol{d}^{\text{obs}}) \propto \widetilde{\mathcal{L}}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\eta}_r)\pi(\boldsymbol{\eta}_r)\pi(\boldsymbol{\eta}_\perp|\boldsymbol{\eta}_r), \quad \boldsymbol{\eta} = A_r\boldsymbol{\eta}_r + A_\perp\boldsymbol{\eta}_\perp$$

⇒ State-of-the-art methods are *gradient-based*.

- $H = \int_{\Xi} \nabla \log \mathcal{L}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\eta}) \left(\nabla \log \mathcal{L}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\eta})\right)^\top \nu(\boldsymbol{\eta})d\boldsymbol{\eta}$
- The reduction relies on the dominant eigenspace of the pencil $(H, C_{\text{prior}}^{-1} = \mathrm{I}_n)$:

$$[u_1 \dots u_r], \quad \lambda_1 \geqslant \dots \geqslant \lambda_n, \quad Hu_i = \lambda_i C_{\text{prior}}^{-1} u_i$$

⇒ Gradients can be expensive/unavailable

Constantine et al., *SIAM JSC*, 2016; **Cui et al.**, *Inv. Prob.*, 2014; **Zahm et al.**, *Math. Comp.*, 2022;

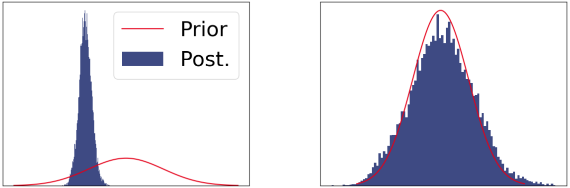
3. Dimension reduction: Gradient-free formulation

Bayesian linear Gaussian case - Linear forward model, Gaussian posterior

Spantini's corollary - The reduction equivalently relies on the minor eigenspace of $(C_{\text{post}}, C_{\text{prior}} = \mathrm{I}_n)$

$$[v_1 \dots v_r], \quad \nu_1 \leqslant \dots \leqslant \nu_n, \quad C_{\text{post}} v_i = \nu_i C_{\text{prior}} v_i$$

New corollary - In the (BLG) case, considering an approximation of the form $\widetilde{\pi}_{\text{post}}$, this projector is also optimal to approximate C_{post}.



(a) Informed direction (b) Non-informed direction

Spantini et al., *SIAM JSC*, 2015

Posterior covariance approximation

- C_{post} is unknown
- Inputs: m samples $\boldsymbol{\eta}^{(i)} \sim \widetilde{\pi}_{\mathrm{post}}$, $\mathcal{L}(\boldsymbol{\eta}^{(i)})$, $\widetilde{\mathcal{L}}({\boldsymbol{\eta}_r}^{(i)})$
- Weights: $\omega^{(i)} = \mathcal{L}(\boldsymbol{\eta}^{(i)})/\widetilde{\mathcal{L}}({\boldsymbol{\eta}_r}^{(i)})$
- Weighted Monte Carlo estimator

$$\widehat{C}_{\mathrm{post}} = \frac{\sum_{i=1}^{m} \omega^{(i)}}{\left(\sum_{i=1}^{m} \omega^{(i)}\right)^2 - \sum_{i=1}^{m} (\omega^{(i)})^2} \sum_{i=1}^{m} \omega^{(i)} (\boldsymbol{\eta}^{(i)} - \overline{\boldsymbol{\eta}})(\boldsymbol{\eta}^{(i)} - \overline{\boldsymbol{\eta}})^\top, \text{ with } \overline{\boldsymbol{\eta}} = \frac{\sum_{i=1}^{m} \omega^{(i)} \boldsymbol{\eta}^{(i)}}{\sum_{i=1}^{m} \omega^{(i)}}$$

- Tempering/shrinkage *etc.* available

El Masri, *PhD thesis*, 2022

Extension to non-linear cases - A bound can be derived from (**Cui and Tong**, *Bern.*, 2022)

Bound of the final approximation quality

The Hellinger distance between the true posterior P and the numerical approximated posterior $P_{\Pi_r}^{(N)}$ is bounded with

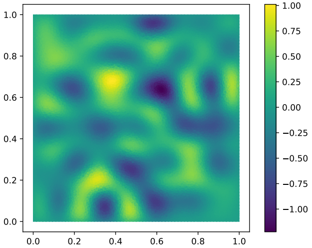
$$D_H(P, P_{\Pi_r}^{(N)}) \leqslant \sqrt{\mathbb{E}_{P_{\Pi_r}}\left(\mathbb{V}\mathrm{ar}_{\pi_{\perp|r}(X_\perp|X_r)}(\sqrt{\omega(X)})\right)} + \sqrt{\frac{2}{N}}\sqrt{\mathbb{E}_{P_{\Pi_r}}\left(\mathbb{V}\mathrm{ar}_{\pi_{\perp|r}(X_\perp|X_r)}(\omega(X))\right)}.$$

Table of contents

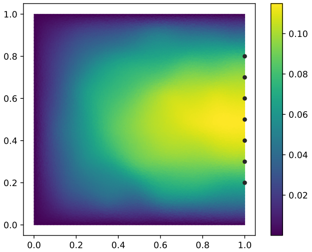
4. Application: Steady-state diffusion equation

$$\begin{cases} -\nabla\left(\kappa(x)\cdot\nabla u(x)\right)=1, & x\in\Omega=[0,1]^2, \\ u(x)=0, & x\in\Gamma_D=\{x, \quad x_1=0\cup x_2=0\cup x_2=1\}, \\ \nabla u(x)\cdot\boldsymbol{n}=0, & x\in\Gamma_N=\{x, \quad x_1=1\}. \end{cases}$$

$\log\kappa\sim\mathcal{N}(0,k)$, k an exponential kernel with correlation length 0.02



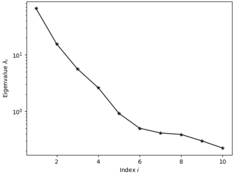
(a) True log-field



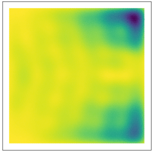
(b) True solution

Constantine et al., *SIAM JSC*, 2016

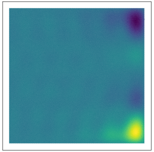
4. Application: Steady-state diffusion equation



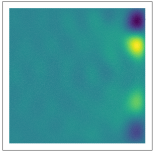
(a) Eigenvalues



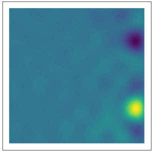
(b) First direction



(c) Second direction



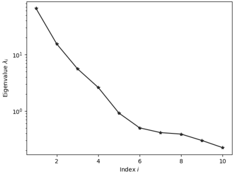
(d) Third direction



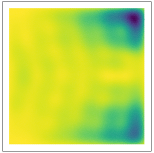
(e) Fourth direction

$(H, C_{\text{prior}}^{-1})$, $n = 1,000$, prior weighted averaged

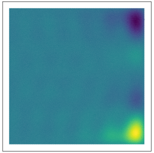
4. Application: Steady-state diffusion equation



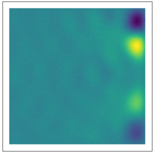
(a) Eigenvalues



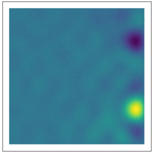
(b) First direction



(c) Second direction

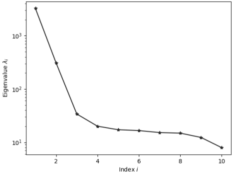


(d) Third direction

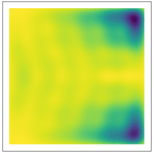


(e) Fourth direction

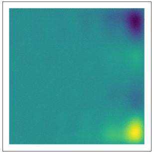
$(H, C_{\text{prior}}^{-1})$, $n = 1,000$, prior weighted averaged



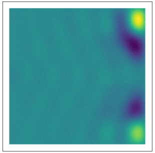
(a) Eigenvalues



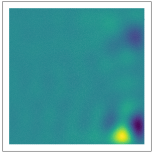
(b) First direction



(c) Second direction



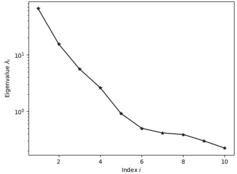
(d) Third direction



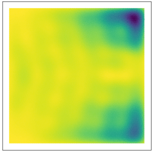
(e) Fourth direction

$(H, C_{\text{prior}}^{-1})$, $n = 1,000$, prior averaged

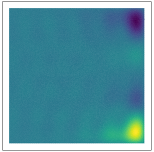
4. Application: Steady-state diffusion equation



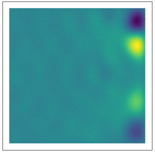
(a) Eigenvalues



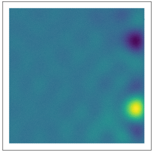
(b) First direction



(c) Second direction

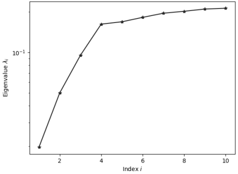


(d) Third direction

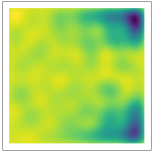


(e) Fourth direction

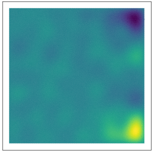
$(H, C_{\text{prior}}^{-1})$, $n = 1,000$, prior weighted averaged



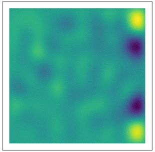
(a) Eigenvalues



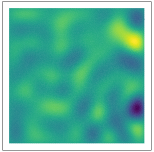
(b) First direction



(c) Second direction



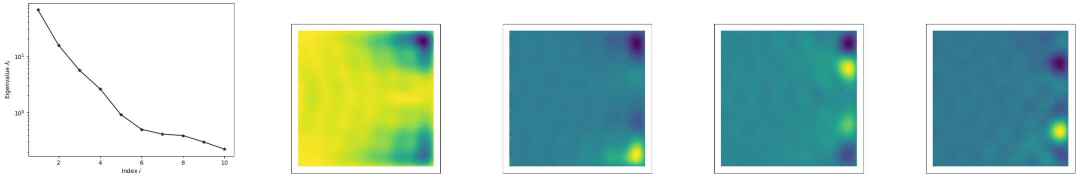
(d) Third direction



(e) Fourth direction

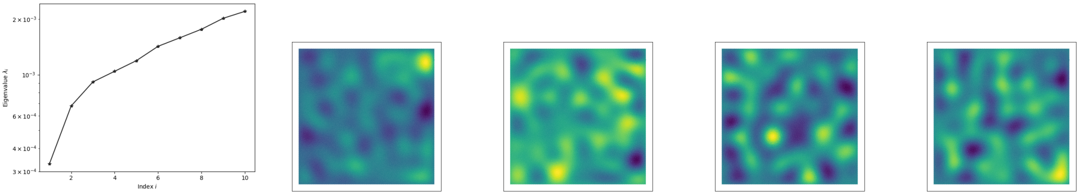
$(\widehat{C}_{\text{post}}, C_{\text{prior}})$, $n = 10,000$

4. Application: Steady-state diffusion equation



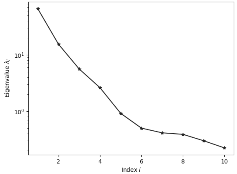
(a) Eigenvalues (b) First direction (c) Second direction (d) Third direction (e) Fourth direction

$(H, C_{\text{prior}}^{-1})$, $n = 1,000$, prior weighted averaged

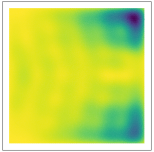


(a) Eigenvalues (b) First direction (c) Second direction (d) Third direction (e) Fourth direction

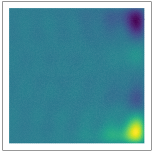
$(\widehat{C}_{\text{post}}, C_{\text{prior}})$, $n = 1,000$



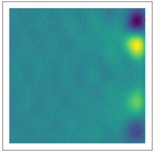
(a) Eigenvalues



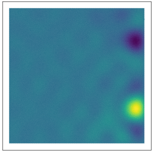
(b) First direction



(c) Second direction

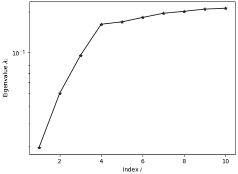


(d) Third direction

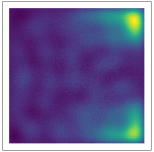


(e) Fourth direction

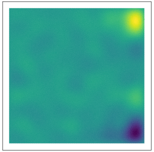
$(H, C_{\text{prior}}^{-1})$, $n = 1,000$, prior weighted averaged



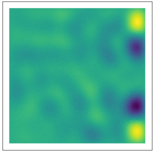
(a) Eigenvalues



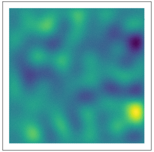
(b) First direction



(c) Second direction



(d) Third direction

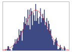
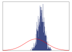
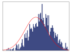
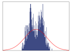
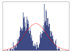
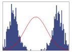


(e) Fourth direction

$(\widehat{C}_{\text{post}}, C_{\text{prior}})$, $n = 1,000$, 8 iterations.

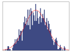
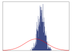
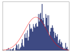
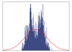
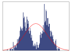
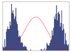
4. Application: Bivariate cases

- Results for bivariate cases: $X = (x_1, x_2)$, $x_1 \sim \pi_{\mathrm{post}}$, $x_2 \sim \mathcal{N}(0,1)$

π_{post}	Illus	λ Grad.-based	Cov.-based
$\mathcal{N}(0,1)$		✓	✓
$\mathcal{N}(0.5,0.1)$		✓	✓
$\mathcal{N}(0.5,1)$		✓	✗
Bi-modal, std$\simeq$ 0.5		✓	✓
Bi-modal, std$\simeq$ 1		✓	✗
Bi-modal, std$\simeq$ 2		✓	✗

4. Application: Bivariate cases

- Results for bivariate cases: $X = (x_1, x_2)$, $x_1 \sim \pi_{\text{post}}$, $x_2 \sim \mathcal{N}(0,1)$
- Correlation coefficient (c.c.) values: $\mathbb{C}\text{or}(X, \log \mathcal{L})$, $\mathbb{C}\text{or}(X^2, \log \mathcal{L})$, Chatterjee coef.

π_{post}	Illus	λ Grad.-based	Cov.-based	lin. c.c.	quad. c.c.	Chatterjee
$\mathcal{N}(0,1)$		✓	✓	N.A.	N.A.	N.A.
$\mathcal{N}(0.5, 0.1)$		✓	✓	✓	✓	✓
$\mathcal{N}(0.5, 1)$		✓	✗	✓	✗	✓
Bi-modal, std$\simeq$ 0.5		✓	✓	✗	✓	✓
Bi-modal, std$\simeq$ 1		✓	✗	✗	✗	✓
Bi-modal, std$\simeq$ 2		✓	✗	✗	✓	✓

Conclusion

- *Objective*: develop numerical methods to help account for field uncertainties in inverse problems
- *Problem*: the parametrization could require a high number of parameters
- *Covariance-Informed Subspace*: adaptive gradient-free dimension reduction
- *Application*: Steady-state diffusion problem $\rightarrow$ from dimension 100 to ~ 4
- Work in progress: using correlation coefficients and partial least squares; application to Global Ozone MOnitoring System (GOMOS)

Thank you !

nadege.polette@minesparis.psl.eu

Keywords: inverse problem, Bayesian inference, MCMC, dimension reduction, gradient free