Acceleration and Reduction Methods for Bayesian Inference

Olivier Le Maître[1]
and Colleagues (many)

[1] Centre de Mathématiques Appliquées, CNRS
Ecole Polytechnique, Palaiseau, France
https://perso.limsi.fr/olm/
olivier.le-maitre@polytechnique.edu

Seminaire interne DéFi

Table of contents

Bayesian inference

Parametric uncertainty

- account for **incomplete knowledge of some model parameters**
- characterize **predictive capabilities** of the uncertain model (GSA)
- deploy uncertainty reduction strategies

Bayes formula

We want to infer a **finite set of parameters** $\boldsymbol{q} \in \mathbb{R}^q$, from

- a set $\mathcal{O} \doteq \{y_i \in \mathbb{R}, i = 1, \dots, M\}$ of measurements,
- a model that predicts the measurements: $\mathbf{U}(\boldsymbol{q}) \in \mathbb{R}^M$

Using the **Bayesian rule** to update our knowledge on $\boldsymbol{q}$:

$$p_{\text{post}}(\boldsymbol{q}|\mathcal{O}) \propto L(\mathcal{O}|\boldsymbol{q})p(\boldsymbol{q}),$$

with

- $L(\mathcal{O}|\boldsymbol{q})$ is the **likelihood** of the measurements,
- $p(\boldsymbol{q})$ is the parameters' **prior**,
- $p_{\text{post}}(\boldsymbol{q}|\mathcal{O})$ is the **posterior**.

Bayesian inference

Likelihood function

Model for the **measurements error** (noise):

$$Y_i = U_i(\boldsymbol{q}) + \epsilon_i, \quad \epsilon_i = N(0, \sigma_i^2),$$

Need a mode for the ϵ,

$$\boldsymbol{\epsilon} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\epsilon}}, \Sigma^2_{\boldsymbol{\epsilon}}\right).$$

Case of unbiased independent Gaussian errors:

$$L(\mathcal{O}|\boldsymbol{q}) \doteq \prod_{i=1}^{M} \exp\left[-\frac{|y_i - U_i(\boldsymbol{q})|^2}{2\sigma_i^2}\right].$$

Note: in practice needs hyper-parameters (*i.e.* noise variance).

Prior of parameter

Reflect our initial knowledge on the parameters.
For instance **Gaussian prior**

$$\boldsymbol{q} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{q}}, \Sigma^2_{\boldsymbol{q}}\right).$$

Current applications

At Inria-CMAP

- inference of complex thermodynamical models (G. Gori, Milan, Utopiae)
- inference of properties for thermal protection systems (A. del Val, VKI, Utopiae)
- inference of reduced combustion model (J. Mateu, EM2C Centrale-Supelec)
- inference of geological properties-tomography (P. Sochala & A. Grenet, Brgm & Ecole des Mines)
- inference of model error (N. Leoni, CMAP Cea)
- Bayesian inversion for oil spills (O. Knio, KAUST)
- ...

Complexity of Bayesian inference

Forward model complexity (see 2)

- Evaluating $L(\mathcal{O}|\boldsymbol{q})$ calls an **evaluation of the model $\boldsymbol{U}(\boldsymbol{q})$**
- **Finding MAP** of $p_{\text{post}}(\boldsymbol{q}|\mathcal{O})$ can be prohibitive
- **Direct sampling** (*e.g.* by MCMC) of the posterior is often impossible

Need for **(forward) model reduction**.

Hyper-parameters (see 3)

- Noise structure (Σ_ϵ^2) are usually unknown a priori
- Specifying **exact** a priori ($\Sigma_{\boldsymbol{q}}^2$) is not suitable
- **Should be learnt or selected from evidence**

Need to "enrich" the inference problem with **hyper-parameters**.

Importance of observation (see 4 & 5)

- **Not all observations** are informative
- Adding observations can **increase computational complexity with no gain**
- Combining observations can result in **better conditioned problems**

Need for objective methods to **select and reduce observations**

Table of contents

Surrogate model for Bayesian Inference

with Didier Lucor & Lionel Mathelin (LIMSI)

[D. Lucor & OLM. ESAIM Proc., 2018]

Standard approach

Inference of $\boldsymbol{q} \in \mathbb{R}^d$ from $\mathcal{O} \doteq \{O_i \in \mathbb{R}, i = 1, \ldots, M\}$ (measurements)
Bayes' formula:

$$p_{\mathrm{post}}(\boldsymbol{q}|\mathcal{O}) \propto L(\mathcal{O}|\boldsymbol{q})p(\boldsymbol{q}),$$

with $p(\boldsymbol{q})$ (prior), $L(\mathcal{O}|\boldsymbol{q})$ (likelihood) and $p_{\mathrm{post}}(\boldsymbol{q}|\mathcal{O})$ (posterior)
Model for the measurement errors:

$$O_i = U_i(\boldsymbol{q}) + \epsilon_i, \quad \epsilon_i = N(0, \sigma_i^2),$$

$U_i(\boldsymbol{q})$ is the model prediction of the i-th measurement
Likelihood becomes:

$$L(\mathcal{O}|\boldsymbol{q}) \doteq \prod_{i=1}^{M} \exp\left[-\frac{|O_i - U_i(\boldsymbol{q})|^2}{2\sigma_i^2}\right].$$

Posterior sampled, for instance using **Markov Chain Monte Carlo (MCMC), rely heavily on multiple evaluations** of

$$\boldsymbol{q} \mapsto \boldsymbol{U}(\boldsymbol{q}) \doteq (U_1 \cdots U_M)(\boldsymbol{q})$$

Surrogate based posterior

Substitute costly model $\boldsymbol{U}$ with a surrogate $\hat{\boldsymbol{U}}$ with inexpensive evaluations.
Classically, surrogate is constructed **off-line**.
The surrogate-based posterior becomes

$$\hat{p}_{\text{post}}(\boldsymbol{q}|\mathcal{O}) \propto \hat{L}(\mathcal{O}|\boldsymbol{q})p(\boldsymbol{q}), \quad \hat{L}(\mathcal{O}|\boldsymbol{q}) \doteq \prod_{i=1}^{M} \exp\left[-\frac{|O_i - \hat{U}_i(\boldsymbol{q})|^2}{2\sigma_i^2}\right].$$

Error estimate [Marzouk, Xiu, Najm, ...]

$$\mathrm{KL}(p_{\text{post}}|\hat{p}_{\text{post}}) \doteq \int\cdots\int \log \frac{p_{\text{post}}(\boldsymbol{q}|\mathcal{O})}{\hat{p}_{\text{post}}(\boldsymbol{q}|\mathcal{O})} p_{\text{post}}(\boldsymbol{q}|\mathcal{O}) d\boldsymbol{q} \leq C(\mathcal{O}) \left(\sum_{i=1}^{M} \|U_i - \hat{U}_i\|_{L_2(p)}^2\right)^{1/2},$$

where

$$\|u\|_{L_2(p)}^2 \doteq \int\cdots\int |u(\boldsymbol{q})|^2 p(\boldsymbol{q}) d\boldsymbol{q}$$

Motivate for surrogate minimizing $\|U_i - \hat{U}_i\|_{L_2(p)}$.
For a priori independent parameters, **PC surrogates** [Marzouk, Najm]

$$U_i(\boldsymbol{q}) \approx \hat{U}_i(\boldsymbol{q}) \doteq \sum_{\alpha=1}^{P} [U_i]_\alpha \Psi_\alpha(\boldsymbol{q}),$$

Supported by the possibly high convergence rate of the approximation.

Surrogate based posterior

Substitute costly model $\boldsymbol{U}$ with a surrogate $\hat{\boldsymbol{U}}$ with inexpensive evaluations.
Classically, surrogate is constructed **off-line**.
The surrogate-based posterior becomes

$$\hat{p}_{\mathrm{post}}(\boldsymbol{q}|\mathcal{O}) \propto \hat{L}(\mathcal{O}|\boldsymbol{q})p(\boldsymbol{q}), \quad \hat{L}(\mathcal{O}|\boldsymbol{q}) \doteq \prod_{i=1}^{M} \exp\left[-\frac{|O_i - \hat{U}_i(\boldsymbol{q})|^2}{2\sigma_i^2}\right].$$

Error estimate `[Marzouk, Xiu, Najm, ...]`

$$\mathrm{KL}(p_{\mathrm{post}}|\hat{p}_{\mathrm{post}}) \doteq \int\cdots\int \log \frac{p_{\mathrm{post}}(\boldsymbol{q}|\mathcal{O})}{\hat{p}_{\mathrm{post}}(\boldsymbol{q}|\mathcal{O})} p_{\mathrm{post}}(\boldsymbol{q}|\mathcal{O}) d\boldsymbol{q} \leq C(\mathcal{O}) \left(\sum_{i=1}^{M} \|U_i - \hat{U}_i\|^2_{L_2(p)}\right)^{1/2},$$

Constant $C(\mathcal{O})$ can be large if the observations are very informative:

$$\mathbb{E}_{p_{\mathrm{post}}}\left\{|U_i - \hat{U}_i|^2\right\} = \int\cdots\int |U_i(\boldsymbol{q}) - \hat{U}_i(\boldsymbol{q})|^2 p_{\mathrm{post}}(\boldsymbol{q}|\mathcal{O}) d\boldsymbol{q}.$$

But the posterior is unknown!

Iterative approach

Basic idea:

- a sequence of polynomial surrogates $\hat{\boldsymbol{U}}^{(k)}(\boldsymbol{q})$ incorporating progressively new observations of $\boldsymbol{U}$
- take new observations of the model to improve the surrogate error (in the posterior norm)

Denote $\mathcal{D} = \{(\boldsymbol{q}^j, \boldsymbol{U}^j, \rho^j), j = 1, \dots, n\}$ the set of collected model observations:

- $\boldsymbol{q}^j$ observation point
- $\boldsymbol{U}^j = \boldsymbol{U}(q^j)$ full model evaluation
- $\rho^j > 0$ trust index

Iterative approach

Basic idea:

- a sequence of polynomial surrogates $\hat{\boldsymbol{U}}^{(k)}(\boldsymbol{q})$ incorporating progressively new observations of $\boldsymbol{U}$
- take new observations of the model to improve the surrogate error (in the posterior norm)

Model construction:

- select a subset $\mathcal{I}^{(k)}$ of model observations indexes
- find the polynomial approximation

$$\boldsymbol{U}(\boldsymbol{q}) \approx \boldsymbol{U}^{(k)}(\boldsymbol{q}) = \sum_{\alpha=1}^{P} [\boldsymbol{U}]_\alpha^{(k)} \Psi_\alpha(\boldsymbol{\eta}^{(k)}(\boldsymbol{q})),$$

- solving a regularized regression problem of type

$$\boldsymbol{u} = \arg \min_{\boldsymbol{\nu} \in \mathbb{R}^P} \sum_{j \in \mathcal{I}} \rho^j \left| U^j - \sum_{\alpha=0}^{P} \Psi_\alpha(\boldsymbol{q}^j) v_\alpha \right|^2 + \lambda \sum_{\alpha=0}^{p} |v_\alpha|.$$

Iterative approach

Basic idea:

- a sequence of polynomial surrogates $\hat{\boldsymbol{U}}^{(k)}(\boldsymbol{q})$ incorporating progressively new observations of $\boldsymbol{U}$
- take new observations of the model to improve the surrogate error (in the posterior norm)

Resampling: (completing the model observations set)

$$\hat{p}^{(k)}_{\text{post}}(\boldsymbol{q}|\mathcal{O}) \propto \exp\left[\sum_{i=1}^{M} -\frac{\left|O_i - \hat{U}_i^{(k)}(\boldsymbol{q})\right|^2}{2\sigma_i^2}\right] p(\boldsymbol{q}).$$

- Draw several independent samples $\boldsymbol{q}^j$ form $\hat{p}^{(k)}_{\text{post}}$
- Compute model prediction $\boldsymbol{U}^j = \boldsymbol{U}(\boldsymbol{q}^j)$
- Define the trust index of the new observation as

$$(\Delta^j)^2 \doteq \sum_{i=1}^{M} \frac{|U_i^j - \hat{U}_i^{(k)}(\boldsymbol{q}^j)|^2}{2\sigma_i^2}, \rho^j \doteq \frac{1}{\max(\epsilon_t, \Delta^j)}.$$

General Iterative Algorithm

ALGORITHM 1: Iterative Procedure for the Construction of the Posterior Fitted Surrogate.

Require: Initial number of observations n_0, number of new observations at each step n_{add}, measurements set $\mathcal{O}$, maximal number of model evaluations $n_{\max}$

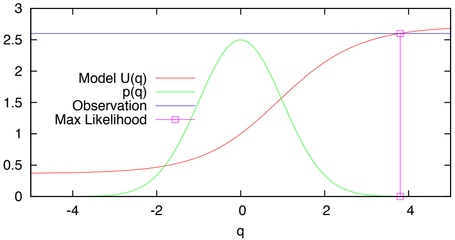
Initialization:
$n = 1$, $\mathcal{D} = \emptyset$ ▷ Initialize the observations set
for $j = 1, \dots n_0$ **do** ▷ Generate the initial observations
  Draw $\boldsymbol{q}^n$ from $p(\boldsymbol{q})$, $\mathcal{D} \leftarrow \mathcal{D} \cup \{(\boldsymbol{q}^n, \boldsymbol{U}(\boldsymbol{q}^n), \rho_0)\}$, $n \leftarrow n + 1$
end for
$k = 0$, construct $\hat{\boldsymbol{U}}^{(0)}$ with $\mathcal{I}^{(0)} = \{1, \dots, n\}$ ▷ Construct initial surrogate
while $n < n_{max}$ **do**
  for $j = 1, \dots n_{add}$ **do**
    Draw $\boldsymbol{q}^n$ from $\hat{p}^{(k)}_{\text{post}}(\boldsymbol{q}|\mathcal{O})$ ▷ Sample surrogate-based posterior
    Compute $\boldsymbol{U}(\boldsymbol{q}^n)$ and observation weight ρ^n from (19) ▷ Set observation
    $\mathcal{D} \leftarrow \mathcal{D} \cup \{(\boldsymbol{q}^n, \boldsymbol{U}(\boldsymbol{q}^n), \rho_0)\}$, $n \leftarrow n + 1$ ▷ Update observation set
  end for
  $k \leftarrow k + 1$
  Define $\mathcal{I}^{(k)}$, construct $\hat{\boldsymbol{U}}^{(k)}$ ▷ Specify observations to use and compute surrogate
end while
Return $\hat{\boldsymbol{U}}^{(k)}$ ▷ Return final surrogate

Elementary 1D problem

Simple one-dimensional test problem

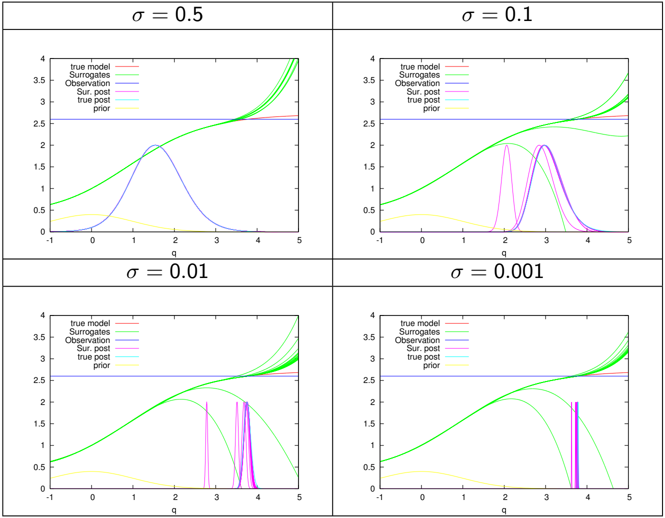
Problem settings

- ✓ $q \in \mathbb{R}^{d=1}$ and non-polynomial model: $U(q) = \exp[\tanh(q/2)]$
- ✓ standard Gaussian prior: $q \sim p(q) = \exp[-q^2/2]/\sqrt{2\pi}$
- ✓ single observation $O = 2.6$, likelihood maximized for $q = 3.8$



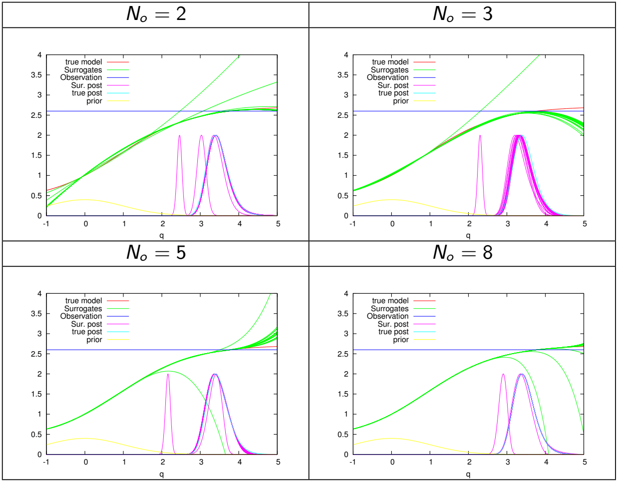
- ✓ for small noise level, $\sigma \ll 1$, prior and posterior are very distant
- ✓ high pol. order N_o required to globally approximate $U(q)$ over few std range

Elementary 1D problem



Elementary 1D problem

Effect of polynomial degree N_o (noise level $\sigma = 0.05$; sampling $|\mathcal{D}^{(k)}|_{k=1\ldots10} = 2N_o$)

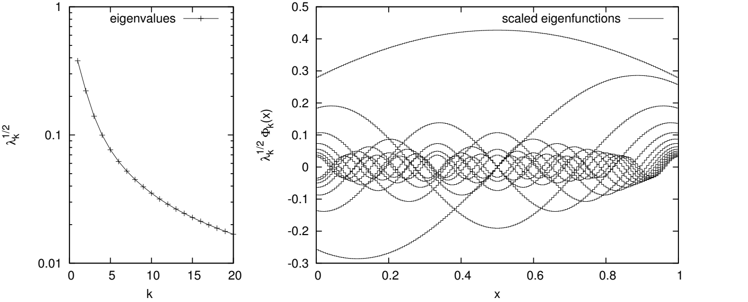


(1D) Elliptic problem

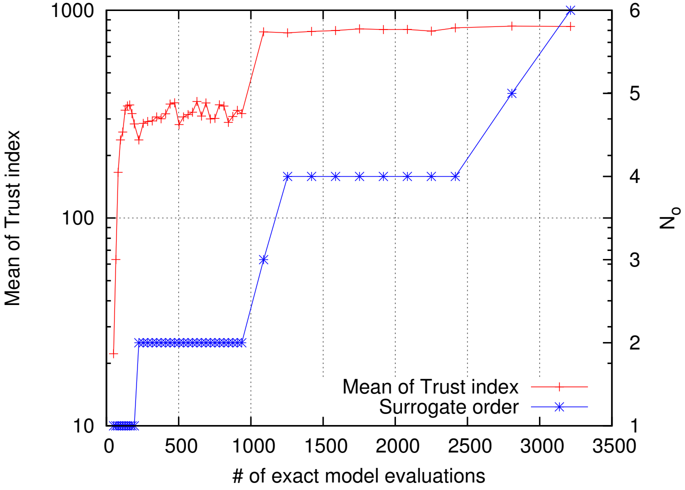
$$\partial\left(\kappa(x)\partial u(x)\right) = -g, \quad \forall x \in]0,1[$$

- Log-normal random field, exponential type covariance
- Retain the first 15 modes: $\boldsymbol{q} \in \mathbb{R}^{15}$

$$\log \kappa(x,\omega) = \sum_{l=1}^{l=15} \sqrt{\lambda_l}\phi_l(x) q_l(\omega), \quad \boldsymbol{q} \sim N(\mathbf{0}, \mathbf{I}).$$



Case of measurements from truth at $\boldsymbol{q} = 0$ and $\sigma = 0.001$



Case of measurements from truth at $\boldsymbol{q} = 0$ and $\sigma = 0.001$

	Iterative Surrogate			Global Surrogate			Error ratio
N_{max} $(\lvert\mathcal{D}\rvert)$	$\epsilon^{(k)}$	${N_o}^{(k)}$	N_{PC}	ϵ^{G}	${N_o}^{G}$	N_{PC}	$\epsilon^{(k)}/\epsilon^{G}$
500 (503)	$3.1\ 10^{-3}$	2	16	$9.4\ 10^{-3}$	4	166	0.33
1000 (1088)	$3.8\ 10^{-4}$	4	166	$6.8\ 10^{-3}$	4	166	0.06
2000 (2084)	$3.7\ 10^{-4}$	4	166	$3.2\ 10^{-3}$	6	406	0.11
2500 (2807)	$2.9\ 10^{-4}$	6	406	$2.7\ 10^{-3}$	6	406	0.11
3000 (3213)	$4.1\ 10^{-4}$	6	406	$2.5\ 10^{-3}$	6	406	0.16

Table 1: Using ${N_o}^{(0)} = 1$, and different N_{max} as indicated. $\sigma = 0.01$.

Case of measurements from truth at $\boldsymbol{q} = 0$ and $\sigma = 0.001$

$N_{max} = 500$	$N_{max} = 2000$	$N_{max} = 3000$

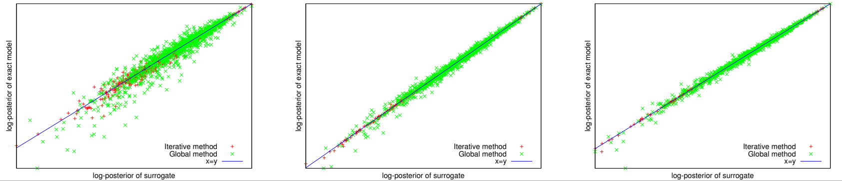


Figure 3: True log-posterior against surrogate log-posteriors values for 1000 sample points drawn from $\hat{p}^{(k)}_{post}$ (Iterative method) and $\hat{p}^{G}_{post}$ (Global method) respectively. Surrogates are constructed with different values of N_{max}, as indicated, and for $\sigma = 0, .01$, $\bar{\boldsymbol{q}} = 0$, ${N_o}^{(0)} = 1$.

Impact of measurement

$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.001$

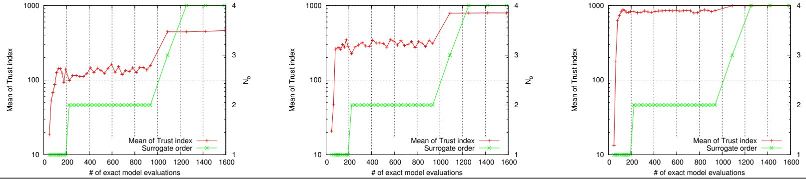


Figure 5: Evolutions of the averaged trust-index for $\bar{\boldsymbol{q}} = 0$, $\mathrm{N}_{\max} = 1500$, ${\mathrm{N}_{\mathrm{o}}}^{(0)} = 1$ and different values for σ as indicated. Also shown are the evolutions of the polynomial order of the successive surrogates (left axis).

Impact of measurement

	$\overline{\Delta} = 0.5$	$\overline{\Delta} = 1.0$	$\overline{\Delta} = 2.0$	N_o	N_{PC}
$\epsilon^{(k)}$	$2.7\ 10^{-5}$	$7.5\ 10^{-6}$	$3.1\ 10^{-6}$	4	166
ϵ^{G}	$2.1\ 10^{-3}$	$7.6\ 10^{-3}$	$2.8\ 10^{-2}$	6	406
$\epsilon^{(k)}/\epsilon^{G}$	$1.3\ 10^{-2}$	$9.9\ 10^{-4}$	$1.1\ 10^{-4}$	-	-

Table 3: Using ${N_o}^{(0)} = 2$, $N_{max} = 1500$, $\sigma = 0.001$.

Impact of measurement

$\overline{\Delta} = 0.5$	$\overline{\Delta} = 1.0$	$\overline{\Delta} = 2.0$

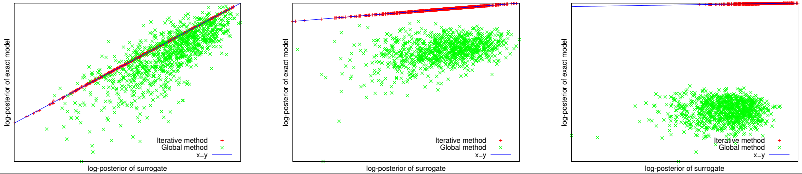


Figure 6: True log-posterior against surrogate log-posteriors values for 1000 sample points drawn from $\hat{p}^{(k)}_{\text{post}}$ (Iterative method) and $\hat{p}^{G}_{\text{post}}$ (Global method) respectively. Case of construction with $\mathrm{N}_{\max} = 1500$, for $\bar{\boldsymbol{q}} = 0$, $\mathrm{N_o}^{(0)} = 1$ and different σ as indicated.

[D. Lucor & OLM. ESAIM Proc., sub.]

Table of contents

Coordinate transformation

Ihab Sraj, Ibrahim Hoteit and Omar Knio (KAUST)

[Sraj, OLM, Hoteit and Knio. Comp. Meth. App. Mech. Eng., 2016]

Example: Inference of a parameter field from Gaussian prior

We want to infer $M \in L_2(\Omega)$, from a Gaussian prior: **centered Gaussian processes with covariance function $\mathcal{C}(x,x')$**.
The prior $M(x)$ **can then be decomposed** in Principal Orthogonal Components (KL decomposition),

$$\mathcal{C}(x,x') = \sum_{k=1}^{\infty} \lambda_k \phi_k(x)\phi_k(x'), \quad M(x) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\Phi_k(x) q_k,$$

where the q_k's are iid standard Gaussian random variables.
Upon truncation of the expansion of M to its K **dominant terms**,

$$M(x) \approx M_K(x) = \sum_{k=1}^{K} \sqrt{\lambda_k}\Phi_k(x) q_k,$$

Inference problem for the stochastic coordinates q_k's:

$$p(\boldsymbol{q}, \sigma^2 | \mathcal{O}) \propto p(\mathcal{O}|\boldsymbol{q}, \sigma^2) p_{\boldsymbol{q}}(\boldsymbol{q}) p_\sigma(\sigma^2),$$

with prior and likelihood

$$p_{\boldsymbol{q}}(\boldsymbol{q}) = \frac{1}{(2\pi)^{K/2}} \exp\left[-\|\boldsymbol{q}\|^2/2\right], \quad p(\mathcal{O}|\boldsymbol{q}, \sigma^2) = \prod_{i=1}^{m} p_\epsilon(O_i - U_i(\boldsymbol{q}), \sigma^2).$$

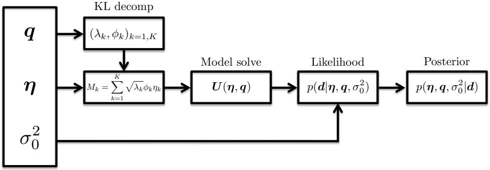
Uncertainty in the covariance function

The selection of the prior **covariance function** affects the inference procedure:
$\Rightarrow$ consider covariance function $\mathcal{C}(\mathbf{h})$ with **hyper-parameters h** having prior $p_h(\mathbf{h})$.
Following this approach, we write

$$M(x,\mathbf{h}) \approx M_K(x,\mathbf{h}) = \sum_{k=1}^{K} \sqrt{\lambda_k(\mathbf{h})}\Phi_k(x,\mathbf{h})q_k,$$

where the q_k's are still i.i.d. standard Gaussian random variables and $(\lambda_k(\mathbf{h}),\Phi_k(\mathbf{h}))$ are the **parametrized dominant proper elements** of $\mathcal{C}(x,x',\mathbf{h})$. It comes

$$p(\boldsymbol{q},\mathbf{h},\sigma^2|\mathcal{O}) \propto p(\mathcal{O}|\boldsymbol{q},\mathbf{h},\sigma^2)p_q(\boldsymbol{q})p_h(\mathbf{h})p_\sigma(\sigma^2).$$



- many KL decomposition
- **many model solves**
- change of coordinate for **h** dependence
- Use of PC surrogate for acceleration

Reference Proper Basis

For any covariance parameters **h**, the elements of the KL expansion are given by

$$\int_\Omega \mathcal{C}(x, x', \mathbf{h})\Phi_k(x', \mathbf{h})dx' = \lambda_k(\mathbf{h})\Phi_k(x, \mathbf{h}), \quad (\Phi_k, \Phi_k)_\Omega = 1.$$

We observe that **$\{\Phi_k(\mathbf{h})\}$ is a CONS of $L_2(\Omega)$**.
It suggests the introduction of a **reference orthonormal basis** $\{\bar{\Phi}_k\}$, defined for a prescribed reference covariance function $\overline{\mathcal{C}}$, and to project $M_k(\mathbf{h})$ into this reference subspace.
Let $\tilde{\Phi}_k(\mathbf{h}) = \sqrt{\lambda_k}(\mathbf{h})\Phi_k(\mathbf{h})$, it comes

$$M_k(x, \mathbf{h}) = \sum_{k=1}^{K} \tilde{\Phi}_k(x, \mathbf{h}) q_k = \sum_{k=1}^{K} \left(\sum_{k'=1}^{\infty} b_{k,k'}(\mathbf{h}) \bar{\Phi}_{k'}(x) \right) q_k, \quad b_{k,k'}(\mathbf{h}) = (\tilde{\Phi}_k(\mathbf{h}), \bar{\Phi}_{k'})_\Omega.$$

For a finite dimensional reference basis (with K modes for simplicity), it comes

$$M_k(x, \mathbf{h}) \approx \tilde{M}_K(x, \mathbf{h}) \doteq \sum_{k=1}^{K} \bar{\Phi}_k(x) \tilde{q}_k(\mathbf{h}), \quad \tilde{\boldsymbol{q}}(\mathbf{h}) = \mathcal{B}(\mathbf{h})\boldsymbol{q}.$$

Change of coordinates

The effect of $\mathbf{h}$ are reflected by a linear change of stochastic coordinates $\boldsymbol{q} \mapsto \tilde{\boldsymbol{q}}(\mathbf{h})$, such that

$$\tilde{\boldsymbol{q}}(\mathbf{h}) = \mathcal{B}(\mathbf{h})\boldsymbol{q} \Rightarrow \Sigma^2(\mathbf{h}) = \mathbb{E}\left[\tilde{\boldsymbol{q}}(\mathbf{h})\tilde{\boldsymbol{q}}^t(\mathbf{h})\right] = \mathcal{B}(\mathbf{h})\mathcal{B}^t(\mathbf{h}).$$

We note that **$\tilde{\boldsymbol{q}}(\mathbf{h})$ is Gaussian with conditional density**

$$p_{\tilde{q}}(\tilde{\boldsymbol{q}}|\boldsymbol{q}) = \frac{1}{\sqrt{2\pi|\Sigma^2(\mathbf{h})|}} \exp\left[-\frac{\tilde{\boldsymbol{q}}^t(\Sigma^2(\mathbf{h}))^{-1}\tilde{\boldsymbol{q}}}{2}\right],$$

where $|\Sigma^2(\mathbf{h})|$ is the determinant of $\Sigma^2(\mathbf{h})$. We shall assume $\Sigma^2(\mathbf{h})$ non-singular a.s.

Regarding the selection of the reference basis:

- select of particular hyper-parameter value: $\overline{\mathcal{C}} = \mathcal{C}(\tilde{\mathbf{h}})$
- use the $\mathbf{h}$-averaged covariance function,

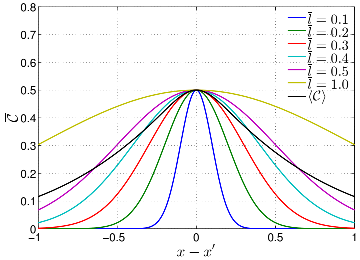
$$\tilde{\mathcal{C}} = \langle \mathcal{C} \rangle = \int \mathcal{C}(\mathbf{h}) p_h(\mathbf{h}) d\mathbf{h}.$$

The latter choice is optimal in terms of representation error (averaged over $\mathbf{h}$).

Example: Gaussian covariance function

Consider $\Omega = [0,1]$ and a Gaussian covariance function with uncertain correlation length:

$$\mathcal{C}(x, x', \mathbf{h} = \{l\}) = \sigma_f^2 \exp\left(-\frac{(x-x')^2}{2l^2}\right), \quad l \sim U[0.1, 1].$$



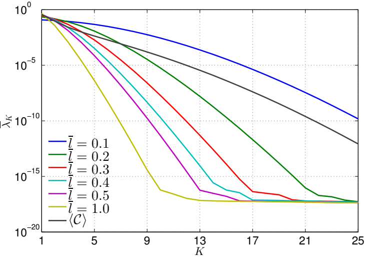


Figure: (Left) Reference covariance functions $\mathcal{C}(\bar{l})$ for different values of $\bar{l}$, as indicated. Also plotted is the $\mathbf{h}$-averaged covariance $\langle \mathcal{C} \rangle$ and (Right) Spectra of the covariance functions shown in the left plot.

Example of Gaussian covariance function

We define the approximation errors:

$$\epsilon_M(K,\mathbf{h}) = \frac{\|M(\mathbf{h}) - \tilde{M}_K(\mathbf{h})\|_{L_2}}{\|M(\mathbf{h})\|_{L_2}}, \quad E_M^2(K) = \int \epsilon_M^2(K,\mathbf{h}) p_q(\mathbf{h}) d\mathbf{h}.$$

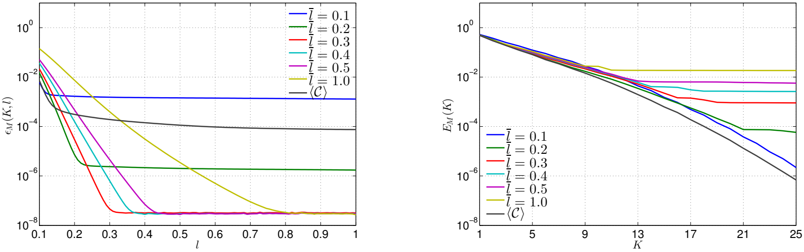


Figure: (Left) Relative error $\epsilon_M(K = 15, l)$ as a function of l. (Right) Error $E_M(K)$ for different reference covariance functions based on selected correlation lengths $\bar{l}$ as indicated. Also plotted are results obtained with $\langle \mathcal{C} \rangle$.

Example of Gaussian covariance function

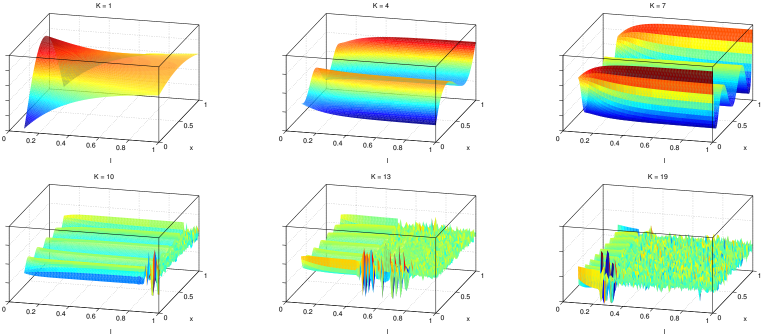


Figure: Dependence of eigen-functions $\phi_k(\mathbf{h})$ with the length-scale hyper-parameter l and selected k as indicated.

Acceleration

Sampling of the posterior $p(\boldsymbol{q}, \mathbf{h}, \sigma^2|\mathcal{O})$ involves **many resolutions of the forward model** to predict the observations $\boldsymbol{U}(\boldsymbol{q}, \mathbf{h})$:

$$p(\boldsymbol{q}, \mathbf{h}, \sigma^2|\mathcal{O}) \propto p(\mathcal{O}|\boldsymbol{q}, \mathbf{h}, \sigma^2) p_q(\boldsymbol{q}) p_h(\mathbf{h}) p_\sigma(\sigma^2), \quad p(\mathcal{O}|\boldsymbol{q}, \mathbf{h}, \sigma^2) = \prod_{i=1}^{m} p_\epsilon(O_i - U_i(\boldsymbol{q}, \mathbf{h}), \sigma^2).$$

To accelerate this step, the use of polynomial surrogates (PC expansions) has been proposed `[Marzouk, Najm, et al]`:

$$\boldsymbol{U}(\boldsymbol{q}, \mathbf{h}) \approx \sum_{\alpha=0}^{P} \boldsymbol{U}_\alpha \Psi_\alpha(\boldsymbol{q}, \mathbf{h}),$$

where the Ψ_α's are orthogonal polynomials and the PC expansion is truncated at some order r.
The PC expansion is computed in an off-line stage and reused by the sampler.
We propose an alternative approach, relying on the coordinate transformation:

$$\boldsymbol{U}(\boldsymbol{q}, \mathbf{h}) \approx \hat{\boldsymbol{U}}(\boldsymbol{\xi}(\boldsymbol{q}, \mathbf{h})) = \sum_{\alpha=0}^{P} \boldsymbol{U}_\alpha \Psi(\boldsymbol{\xi}(\boldsymbol{q}, \mathbf{h})),$$

where the random vector $\boldsymbol{\xi} \in \mathbb{R}^K$.

PC surrogate

Recall that the transformed coordinates $\tilde{\boldsymbol{q}}$ have for conditional density

$$p_{\tilde{q}}(\tilde{\boldsymbol{q}}|\boldsymbol{q}) = \frac{1}{\sqrt{2\pi|\Sigma^2(\mathbf{h})|}} \exp\left[-\frac{\tilde{\boldsymbol{q}}^t(\Sigma^2(\mathbf{h}))^{-1}\tilde{\boldsymbol{q}}}{2}\right].$$

For $\tilde{\mathcal{C}} = \langle\mathcal{C}\rangle$, it can be shown that

$$\int\cdots\int p_{\tilde{q}}(\tilde{\boldsymbol{q}}|\boldsymbol{q})p_q(\mathbf{h})d\mathbf{h} = \frac{1}{\sqrt{2\pi|\Lambda^2|}} \exp\left[-\frac{\tilde{\boldsymbol{q}}^t(\Lambda^2)^{-1}\tilde{\boldsymbol{q}}}{2}\right], \quad \Lambda^2 = \mathrm{diag}\,(\tilde{\lambda}_1 \cdots \tilde{\lambda}_K)$$

It suggests approximating $\tilde{\boldsymbol{q}} \mapsto \boldsymbol{u}(\tilde{\boldsymbol{q}})$ using the reference Gaussian field

$$\tilde{M}_K^{\mathrm{PC}}(\boldsymbol{\xi}) \doteq \sum_{k=1}^{K} \sqrt{\overline{\lambda}_k}\overline{\phi}_k \xi_k, \quad \boldsymbol{\xi} \mapsto \hat{\boldsymbol{u}}(\boldsymbol{\xi}) \approx \sum_{\alpha=0}^{P} \hat{\boldsymbol{u}}_\alpha \Psi_\alpha(\boldsymbol{\xi}),$$

where the **ξ_k's are independent standard Gaussian random variables**. Then

$$\boldsymbol{U}(\boldsymbol{q},\mathbf{h}) \approx \sum_{\alpha=0}^{P} \hat{\boldsymbol{U}}_\alpha \Psi_\alpha(\boldsymbol{\xi}(\boldsymbol{q},\mathbf{h})), \quad \boldsymbol{\xi}(\boldsymbol{q},\mathbf{h}) = \tilde{\mathcal{B}}(\mathbf{h})\boldsymbol{q}, \quad \tilde{\mathcal{B}}_{kl}(\mathbf{h}) = \begin{cases} \dfrac{\mathcal{B}_{kl}(\mathbf{h})}{\sqrt{\overline{\lambda}_k}}, & \overline{\lambda}_k/\overline{\lambda}_1 > \kappa, \\ 0, & \text{otherwise.} \end{cases}$$

for some small $\kappa > 0$.

Conditioning of the coordinate transformation

The PC surrogate is constructed assuming $\boldsymbol{\xi} \sim N(0, \mathbb{I})$; it is subsequently used with $\boldsymbol{\xi}(\boldsymbol{q}, \mathbf{h}) = \tilde{\mathcal{B}}(\mathbf{h})\boldsymbol{q}$. Let $\Sigma^2_{\boldsymbol{\xi}}(\mathbf{h}) = \tilde{\mathcal{B}}(\mathbf{h})^t \tilde{\mathcal{B}}(\mathbf{h})$ and denote $\beta_{\max}(\mathbf{h})$ the largest eigen-value of $\Sigma^2_{\boldsymbol{\xi}}(\mathbf{h})$. It measures the **local stretching of the coordinate transformation**.

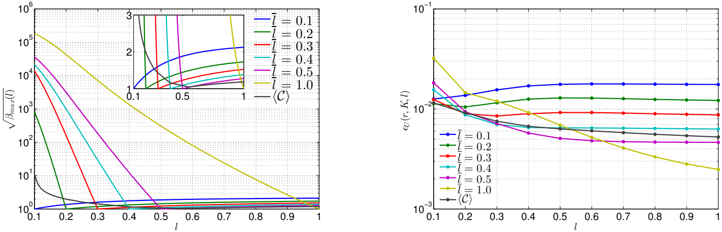


Figure: Left: max stretching $\sqrt{\beta_{max}(l)}$ depending on the selected reference covariance function. Right: corresponding L_2 error of the PC surrogates for $K = 10$ and PC degree $r = 10$.

Sampling flow-chart

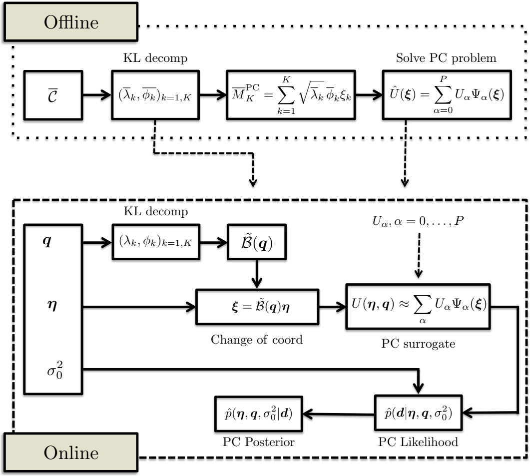


Figure: Off-line step (surrogate construction) of the accelerated MCMC sampler and on-line step of the PC surrogate based evaluation of the posterior.

Example: 1-D diffusion problem

Consider the diffusion problem for $x \in (0,1)$ and $t \in [0, T_f]$, given by

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\nu \frac{\partial u}{\partial x}\right), \quad \nu = \nu_0 + \exp(M),$$

with IC $u = 0$ and BCs $u(x=0,t) = -1$, $u(x=1,t) = 1$ and M is a (centered) Gaussian process with the previous uncertain Gaussian covariance function $\mathcal{C}(\mathbf{h} = \{l\})$.

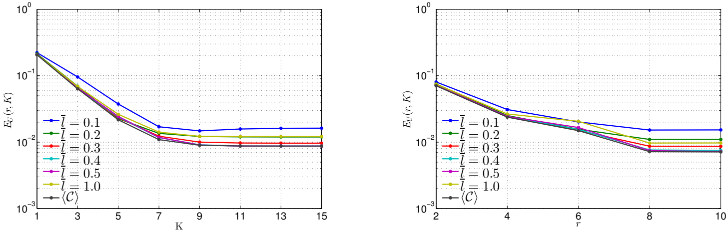


Figure: Global error $E_U(r,K)$ of the PC approximation $\hat{U}$ of the diffusion model problem solution. The left plot shows the dependence of the error with K using a PC order $r = 10$, while the right plot is for different r and $K = 10$. The curves correspond to different definitions of the reference covariance function $\overline{\mathcal{C}}$: $\mathcal{C}(\bar{l})$ with $\bar{l}$ as indicated or the **h**-average covariance function $\langle \mathcal{C} \rangle$.

Test problem

Inference for "true" log-diffusivity fields:

- Sinusoidal profile: $M^{\sin}(x) = \sin(\pi x)$,
- Step function: $M^{\rm step}(x) = \begin{cases} -1/2, & x < 0.5 \\ 1/2, & x \geq 0.5 \end{cases}$,
- Random profile: $M^{\rm ran}(x)$ drawn at random from $\mathcal{GP}(0, \mathcal{C})$ where $\mathcal{C}$ is the Gaussian covariance with length-scale $l = 0.25$ and variance $\sigma_f^2 = 0.65$.

Observations are measurements of $U(x,t)$ at several locations in space and time, corrupted with i.i.d. $\epsilon_i \sim N(0, \sigma_\epsilon^2 = 0.01)$.
For the prior, we use $M \sim \mathcal{GP}(0, \mathcal{C}(\mathbf{h}))$, with Gaussian covariance $\mathcal{C}(\mathbf{h})$ and hyper-parameter $\mathbf{h} = \{l, \sigma_f^2\}$:

- $l \sim U[0.1, 1]$,
- $\sigma_f^2 \sim Inv\Gamma(\alpha, \beta)$, with mean 0.5 and variance 0.25.

Inference **without** covariance Hyper-parameters

We set $l = 0.5$ and $\sigma_f^2 = 0.5$. Also $K = 10$ and $r = 10$.

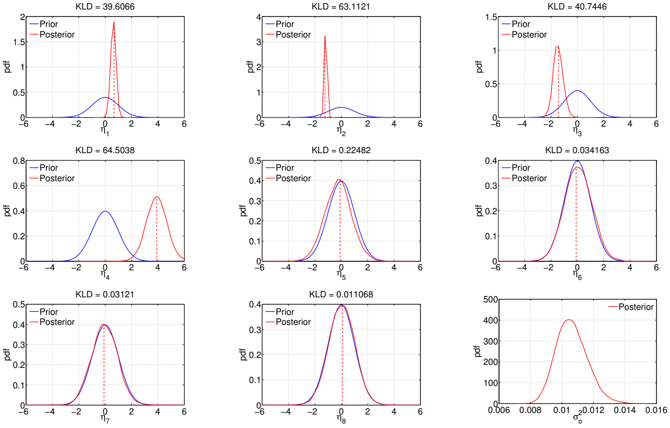


Figure: Comparison of priors and marginals posterior of the first 8 KL coordinates q_k for the **inference of** $M^{\sin}$ without using fixed Gaussian covariance with $l = 0.5$ **and** $\sigma_f^2 = 0.5$. The Kullback-Leibler Divergence (KLD) between the priors and marginal posteriors are also indicated on top of each plot. The posterior of the noise hyper-parameter σ^2 is indicated.

Inference with Hyper-parameters

$K = 10$ and $r = 10$.

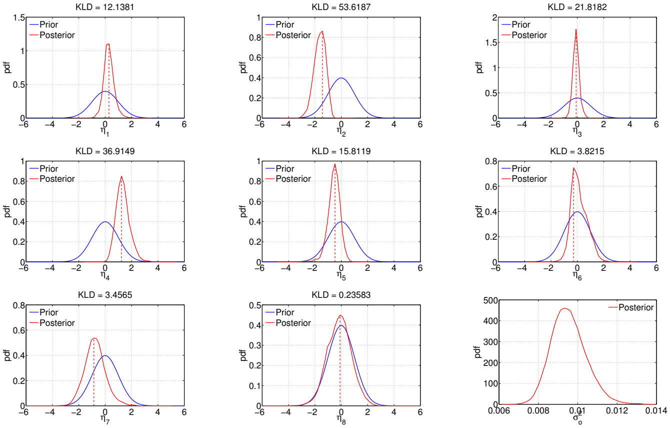


Figure: Comparison of priors and marginals posterior of the first 8 KL coordinates q_k for the **inference of $M^{\sin}$** using covariance hyper-parameters, coordinate transformation and PC surrogate. The Kullback-Leibler Divergence (KLD) between the priors and marginal posteriors are also indicated on top of each plot. The posterior of the noise hyper-parameter σ^2 is indicated.

Inference: comparison of inferred field

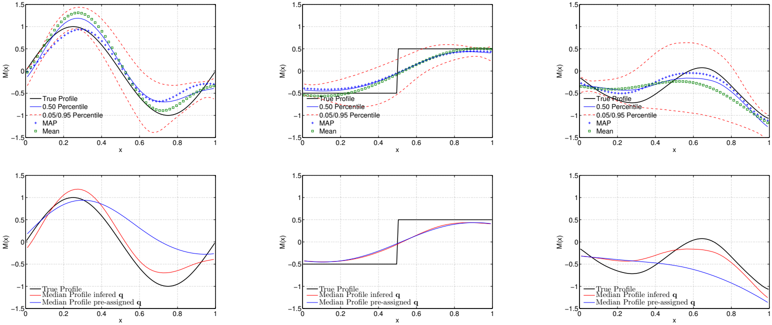


Figure: Comparison of inferred log-diffusivity profile: fixed covariance versus covariance and hyper-parameters.

Inference of Hyper-parameters

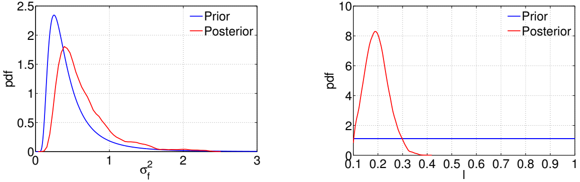


Figure: Posterior pdfs of sinusoidal log-diffusivity profile hyper-parameters.

[Sraj, OLM, Hoteit and Knio. Comp. Meth. App. Mech. Eng., 2016]

Table of contents

Selection of observations

with Maria Navarro, Ibrahim Hoteit, Omar Knio (KAUST)
Kyle Mandli (Columbia) and David George (USGS)

[Navarro, OLM, Mandli, George, Hoteit and Knio. Comp. Geosciences, 2018.]

Debris flow model

- Flow of debris (mud, gravels, small rocks, ...)
- Empirical / Phenomenological models
- Parameter calibration on experiments at USGS

Governing equations `GeoClaw`

$$
\begin{aligned}
&\frac{\partial h}{\partial t} + \frac{\partial (hu)}{\partial x} + \frac{\partial (hv)}{\partial y} = \varphi_1,\\
&\frac{\partial (hu)}{\partial t} + \frac{\partial}{\partial x}(hu^2) + \kappa \frac{\partial}{\partial y}(0.5 g_z h^2) + \frac{\partial (huv)}{\partial y} + \frac{h(1-\kappa)}{\rho}\frac{\partial p_b}{\partial x} = \varphi_2,\\
&\frac{\partial (hv)}{\partial t} + \frac{\partial (huv)}{\partial x} + \frac{\partial}{\partial y}(hv^2) + \kappa \frac{\partial}{\partial y}(0.5 g_z h^2) + \frac{h(1-\kappa)}{\rho}\frac{\partial p_b}{\partial y} = \varphi_3,\\
&\frac{\partial (hm)}{\partial t} + \frac{\partial (hum)}{\partial x} + \frac{\partial (hvm)}{\partial y} = \varphi_4,\\
&\frac{\partial p_b}{\partial t} - \chi u \frac{\partial h}{\partial x} + \chi \frac{\partial (hu)}{\partial x} + u \frac{\partial p_b}{\partial x} - \chi v \frac{\partial h}{\partial y} + \chi \frac{\partial hu}{\partial y} + v \frac{\partial p_b}{\partial y} = \varphi_5.
\end{aligned}
$$

Debris flow model

- Flow of debris (mud, gravels, small rocks, ...)
- Empirical / Phenomenological models
- Parameter calibration on experiments at USGS

Non-linear source terms [Iverson & George, 2014]

$$\varphi_1 = \frac{(\rho - \rho_f)}{\rho} \frac{-2k}{h\mu}(p_b - \rho_f g_z h),$$

$$\varphi_2 = h g_x + u \frac{(\rho - \rho_f)}{\rho} \frac{-2k}{h\mu}(p_b - \rho_f g_z h) - \frac{(\tau_{s,x} + \tau_{f,x})}{\rho},$$

$$\varphi_3 = h g_y + v \frac{(\rho - \rho_f)}{\rho} \frac{-2k}{h\mu}(p_b - \rho_f g_z h) - \frac{(\tau_{s,y} + \tau_{f,y})}{\rho},$$

$$\varphi_4 = \frac{2k}{hu}(p_b - \rho_f g_z h) m \frac{\rho_f}{\rho},$$

$$\varphi_5 = \zeta \frac{-2k}{h\mu}(p_b - \rho_f g_z h) - \frac{3}{\alpha h}\|\boldsymbol{u}\| \tan(\psi),$$

where

$$\zeta = \frac{3}{2\alpha h} + \frac{g_z \rho_f (\rho - \rho_f)}{4\rho}, \quad \alpha = \frac{a}{m(\rho g_z h - p_b + \sigma_0)}.$$

Debris flow model

- Flow of debris (mud, gravels, small rocks, ...)
- Empirical / Phenomenological models
- Parameter calibration on experiments at USGS

Inference of model parameters [Iverson & George, 2014]

- static critical-state solid volume fraction (m_{crit})
- initial hydraulic permeability k_0
- pure-fluid viscosity μ
- steady friction contact angle ϕ
- compressibility constant a.

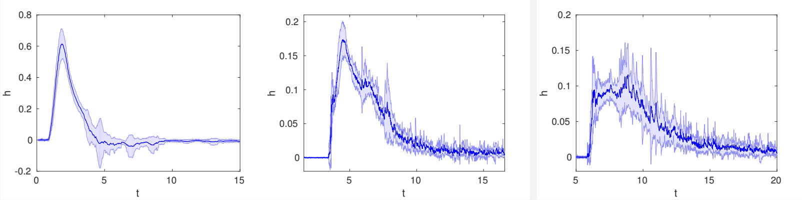
Debris flow model

- Flow of debris (mud, gravels, small rocks, ...)
- Empirical / Phenomenological models
- Parameter calibration on experiments at USGS

Inference of model parameters [Iverson & George, 2014]

- static critical-state solid volume fraction (m_{crit})
- initial hydraulic permeability k_0
- pure-fluid viscosity μ
- steady friction contact angle ϕ
- compressibility constant a.

Gate release experiments: available measurements



Calibration parameters and surrogate model

A priori range of model parameters

$$m_{\text{crit}} \sim \mathscr{U}[0.62, 0.66], \quad k_0 \sim \mathscr{U}_{\log}[10^{-9}, 10^{-8}],$$
$$\mu \sim \mathscr{U}_{\log}[0.005, 0.05], \quad \phi \sim \mathscr{U}[0.62, 0.66], \quad a \sim \mathscr{U}[0.01, 0.05].$$

Parameters considered independent: introduction of canonical random variables

$$m_{crit}(\xi_1), \quad k_0(\xi_2), \quad \mu(\xi_3), \quad \phi(\xi_4), \quad a(\xi_5),$$

where $\boldsymbol{\xi} = (\xi_1 \cdots \xi_5) \sim U[0,1]^5$.

Polynomial Chaos expansions

$U(\boldsymbol{\xi}) \in L_2(\Xi)$ has a PC expansion of the form

$$U(\boldsymbol{\xi}) \approx \sum_{\boldsymbol{\alpha} \in \mathcal{A}} u_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}),$$

- $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \psi_{\alpha_1}(\xi_1) \times \cdots \times \psi_{\alpha_5}(\xi_5)$, where $\{\phi_\alpha, \alpha = 1, 2, \dots\}$ is the set of orthonormal Legendre polynomials
- $\mathcal{A}$ is a multi-index set, we used $\mathcal{A} = \{\boldsymbol{\alpha} \in \mathbb{N}^5, |\boldsymbol{\alpha}| \leq p\}$

Calibration parameters and surrogate model

A priori range of model parameters

$$m_{\text{crit}} \sim \mathscr{U}[0.62, 0.66], \quad k_0 \sim \mathscr{U}_{\log}[10^{-9}, 10^{-8}],$$
$$\mu \sim \mathscr{U}_{\log}[0.005, 0.05], \quad \phi \sim \mathscr{U}[0.62, 0.66], \quad a \sim \mathscr{U}[0.01, 0.05].$$

Parameters considered independent: introduction of canonical random variables

$$m_{crit}(\xi_1), \quad k_0(\xi_2), \quad \mu(\xi_3), \quad \phi(\xi_4), \quad a(\xi_5),$$

where $\boldsymbol{\xi} = (\xi_1 \cdots \xi_5) \sim U[0,1]^5$.

Pre-conditioning the debris height

Makes use of pre-conditioners, defined from

Scaling factor	Symbol	Definition
Arrival time	$t_{\text{arr}}(\boldsymbol{\xi})$	First time $h(t, \boldsymbol{\xi})$ exceeds $\varepsilon \ll 1$
Height at maximum	$h_{\max}(\boldsymbol{\xi})$	$h_{\max}(\boldsymbol{\xi}) = \max_t h(t, \boldsymbol{\xi})$
Time at maximum	$t_{\max}(\boldsymbol{\xi})$	$t_{\max}(\boldsymbol{\xi}) = \arg\max_t h(t, \boldsymbol{\xi})$
Decay time	$t_{\text{dec}}(\boldsymbol{\xi})$	$t_{\text{dec}}(\boldsymbol{\xi}) > t_{\max}(\boldsymbol{\xi})$ such that $h(t_{\text{dec}}, \boldsymbol{\xi}) = 0.4 h_{\max}(\boldsymbol{\xi})$

Table 1 Definition of the scaling factors for the preconditioning.

Calibration parameters and surrogate model

A priori range of model parameters

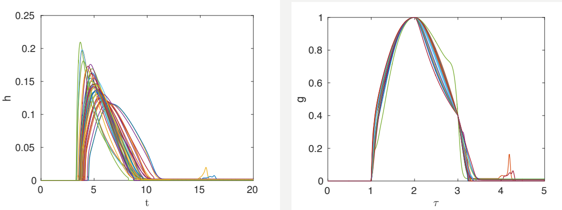
$$m_{\text{crit}} \sim \mathscr{U}[0.62, 0.66], \quad k_0 \sim \mathscr{U}_{\log}[10^{-9}, 10^{-8}],$$
$$\mu \sim \mathscr{U}_{\log}[0.005, 0.05], \quad \phi \sim \mathscr{U}[0.62, 0.66], \quad a \sim \mathscr{U}[0.01, 0.05].$$

Parameters considered independent: introduction of canonical random variables

$$m_{crit}(\xi_1), \quad k_0(\xi_2), \quad \mu(\xi_3), \quad \phi(\xi_4), \quad a(\xi_5),$$

where $\boldsymbol{\xi} = (\xi_1 \cdots \xi_5) \sim U[0,1]^5$.

Transformed predictions



(a) Unscaled realizations (b) Scaled realizations

Calibration parameters and surrogate model

A priori range of model parameters

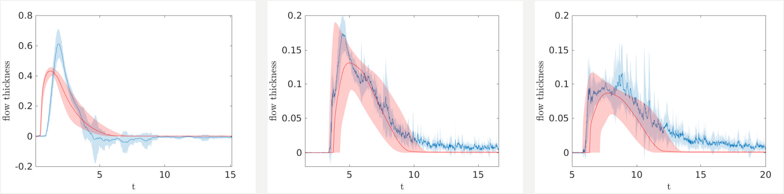
$$m_{\text{crit}} \sim \mathscr{U}[0.62, 0.66], \quad k_0 \sim \mathscr{U}_{\log}[10^{-9}, 10^{-8}],$$
$$\mu \sim \mathscr{U}_{\log}[0.005, 0.05], \quad \phi \sim \mathscr{U}[0.62, 0.66], \quad a \sim \mathscr{U}[0.01, 0.05].$$

Parameters considered independent: introduction of canonical random variables

$$m_{crit}(\xi_1), \quad k_0(\xi_2), \quad \mu(\xi_3), \quad \phi(\xi_4), \quad a(\xi_5),$$

where $\boldsymbol{\xi} = (\xi_1 \cdots \xi_5) \sim U[0,1]^5$.

A priori range



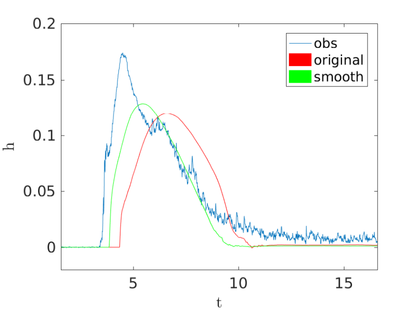
(a) $x = 2$ m (b) $x = 32$ m (c) $x = 66$ m

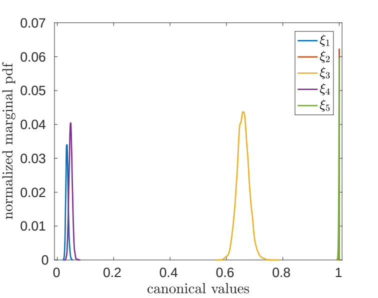
Independent measurement errors

Naive model: Gaussian likelihood

$$\mathcal{L}(\boldsymbol{d}|\boldsymbol{\xi}) = \prod_{i=1}^{m_d} \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp\left[-\frac{(\bar{h}_i - \widehat{h}_i(\boldsymbol{\xi}))^2}{2\sigma_i^2}\right]$$

Independent / uncorrelated **"measurement noise"**

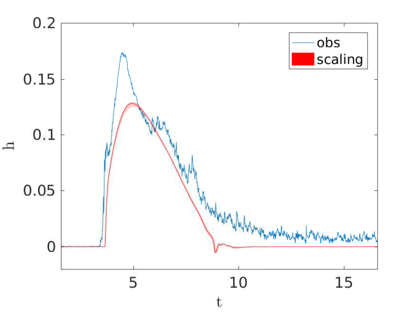


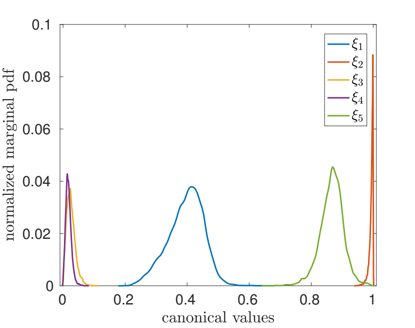


Appreciating inference quality

Trying to fit "important characteristics"

$$\ln(\mathcal{L}(\boldsymbol{d}|\boldsymbol{\xi})) \propto -\left(\frac{t_{\text{arr}} - \widehat{t}_{\text{arr}}(\boldsymbol{\xi})}{2\sigma_{arr}}\right)^2 - \left(\frac{t_{\max} - \widehat{t}_{\max}(\boldsymbol{\xi})}{2\sigma_{t_{\max}}}\right)^2 - \left(\frac{t_{\text{dec}} - \widehat{t}_{\text{dec}}(\boldsymbol{\xi})}{2\sigma_{\text{dec}}}\right)^2 - \left(\frac{h_{\max} - \widehat{h}_{\max}(\boldsymbol{\xi})}{2\sigma_{h_{\max}}}\right)^2.$$



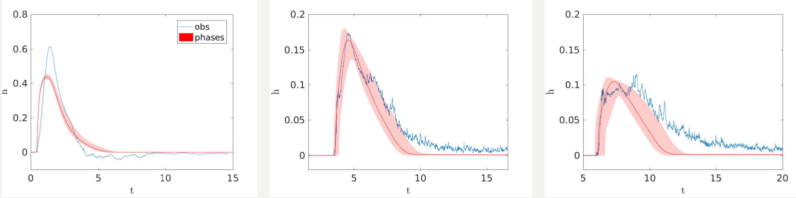


Limits of the model - experimental issues

With feedback from experimentalist

Measurements were synchronized:

$$\ln(\mathcal{L}(\boldsymbol{d}|\boldsymbol{\xi})) \propto -\left(\frac{T_{\text{grw}} - \widehat{T_{\text{grw}}}(\boldsymbol{\xi})}{2\sigma_{T_{\text{grw}}}}\right)^2 - \left(\frac{T_{\text{dec}} - \widehat{T_{\text{dec}}}(\boldsymbol{\xi})}{2\sigma_{T_{\text{dec}}}}\right)^2 - \left(\frac{h_{\max} - \widehat{h}_{\max}(\boldsymbol{\xi})}{2\sigma_{h_{\max}}}\right)^2,$$



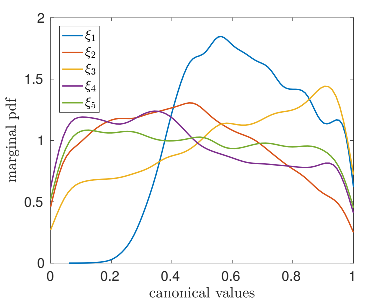
(a) $x = 2m$ (b) $x = 32m$ (c) $x = 66m$

Limits of the model - experimental issues

With feedback from experimentalist

Measurements were synchronized:

$$\ln(\mathcal{L}(\boldsymbol{d}|\boldsymbol{\xi})) \propto -\left(\frac{T_{\mathrm{grw}} - \widehat{T_{\mathrm{grw}}}(\boldsymbol{\xi})}{2\sigma_{T_{\mathrm{grw}}}}\right)^2 - \left(\frac{T_{\mathrm{dec}} - \widehat{T_{\mathrm{dec}}}(\boldsymbol{\xi})}{2\sigma_{T_{\mathrm{dec}}}}\right)^2 - \left(\frac{h_{\max} - \widehat{h}_{\max}(\boldsymbol{\xi})}{2\sigma_{h_{\max}}}\right)^2,$$



Take-away

What did we learn?

- **Experimental data may be biased**
- Raw measurements, or complete description of their treatments, are important
- Using all the available data may be counterproductive (yes!)
- If the model is poor, we should focus on basic features of interest, and not insist on obtaining global agreement
- **Models of model error are more robust and easier to propose & test for simple features**

How to select / reduce the experimental data to facilitate the inference problem?

[Navarro, OLM, Mandli, George, Hoteit and Knio. Comp. Geosciences, 2018.]

Table of contents

Optimal Observations Reduction

Loic Giraldi, Ibrahim Hoteit and Omar Knio (KAUST)

[Giraldi, OLM, Hoteit and Knio. Comp. Stat. & Data Anal., 2018.]

Optimal Observations Reduction

Motivation

Bayesian inference in the case of overabundant data

- Weather forecasting
- Seismic wave inversion

Goal

Compute an optimal approximation

$$\min_{V} \mathscr{L}\left(P(Q \mid Y = y), P(Q \mid W = V^T y)\right)$$

- $\mathscr{L}$ a loss function
- n (random) observations $Y = (Y_i)_{i=1}^{n}$
- q parameters $Q = (Q_i)_{i=1}^{\mathrm{Nq}}$, $\mathrm{Nq} \ll n$
- r dimensional reduced space $V \in \mathbb{R}^{n \times r}$, $r \ll n$

[Giraldi, OLM, Hoteit and Knio. Comp. Stat. & Data An., sub.]

Linear Gaussian models

Gaussian model

$$Y = BQ + E,$$

- Observations: $Y \sim \mathcal{N}(m_Y, C_Y)$ with values in $\mathbb{R}^n$
- Parameter of interest: $Q \sim \mathcal{N}(m_Q, C_Q)$ with values in $\mathbb{R}^{\mathrm{Nq}}$
- Noise: $E \sim \mathcal{N}(m_E, C_E)$ with values in $\mathbb{R}^n$
- Design matrix: $B \in \mathbb{R}^{n \times \mathrm{Nq}}$
- Forward model: $A(Q) = BQ \sim \mathcal{N}(m_A, C_A)$, and $C_{AQ} = \mathrm{Cov}(A(Q), Q)$

Reduced model

$$W = V^T BQ + V^T E,$$

- Reduced observations: $W \sim \mathcal{N}(m_W, C_W)$ with values in $\mathbb{R}^r$
- Reduced space: $V \in \mathbb{R}^{n \times r}$

Posterior distributions

knowing the realization (a particular measurement) y of Y

Unreduced case

The posterior distribution is $P(Q \mid Y = y) \sim \mathcal{N}(m_\star, C_\star)$ where

$$C_\star = C_Q \left(C_Q + C_{AQ}^T C_E^{-1} C_{AQ} \right)^{-1} C_Q,$$
$$m_\star = C_{AQ}^T C_Y^{-1}(y - m_E) + C_\star C_Q^{-1} m_Q.$$

Reduced model

The posterior distribution is $P(Q \mid W = V^T y) \sim \mathcal{N}(m_V, C_V)$ where

$$C_V = C_Q \left(C_Q + C_{AQ}^T V \left(V^T C_E V \right)^{-1} V^T C_{AQ} \right)^{-1} C_Q,$$
$$m_V = C_{AQ}^T V (V^T C_Y V)^{-1} V^T (y - m_E) + C_V C_Q^{-1} m_Q.$$

Invariance property

Proposition (Invariance property)

For all invertible matrices $M \in \mathbb{R}_*^{r\times r}$, we have

$$m_{VM} = m_V \quad \text{and} \quad C_{VM} = C_V.$$

- Posterior distribution invariant under rescaling, rotation or permutation of the observations
- Newton method can not be directly used
- range(V) is more important than V
- Use of a Riemannian trust region algorithm on the Grassmann manifolds Gr(r, n), the set of r-dimensional subspaces of $\mathbb{R}^n$ (see Absil et al. 2007, Manopt and Pymanopt libraries)

Kullback-Leibler based loss functions

Kullback-Leibler divergence

Given two distributions $P(Z_0)$ and $P(Z_1)$ with densities f_{Z_0} and f_{Z_1},

$$\mathrm{D}_{\mathrm{KL}}\left(P(Z_0) \parallel P(Z_1)\right) = \mathbb{E}_{Z_0}\left(\log \frac{f_{Z_0}}{f_{Z_1}}\right).$$

- Quantify the "information lost when $[P(Z_1)]$ is used to approximate $[P(Z_0)]$" (Burnham and Anderson, 2003)
- Positive and null iff $P(Z_0) = P(Z_1)$
- Asymmetric quantity

Kullback-Leibler based loss functions

Kullback-Leibler divergence minimization

$$\min_{[V]\in \mathrm{Gr}(r,n)} \mathrm{D}_{\mathrm{KL}}\left(P(Q \mid Y = y) \,\|\, P(Q \mid W = V^T y)\right)$$

- Closed form of the functional available
- A solution to the optimization problem exists
- **A posteriori reduction** (measurement available)

Expected Kullback-Leibler divergence minimization

$$\min_{[V]\in \mathrm{Gr}(r,n)} \mathbb{E}_Y\left(\mathrm{D}_{\mathrm{KL}}\left(P(Q \mid Y) \,\|\, P(Q \mid W = V^T Y)\right)\right)$$

- Closed form of the functional available
- A solution to the optimization problem exists
- **A priori reduction**

Information-based loss function

Given random variables Z, Z_0, and Z_1,

Entropy

With $Z \sim P(Z)$,

$$\mathrm{H}(Z) = \mathbb{E}_Z(-\log(f_Z(Z))).$$

- Amount of information contained by $P(Z)$

Mutual information

With $Z_0 \sim P(Z_0)$ and $Z_1 \sim P(Z_1)$,

$$\mathcal{I}(Z_0, Z_1) = \mathrm{H}(Z_0) + \mathrm{H}(Z_1) - \mathrm{H}(Z_0, Z_1),$$

- Amount of information that $P(Z_0)$ contains about $P(Z_1)$
- Symmetric quantity

Mutual information maximization

Theorem (Mutual information maximization)

We have

$$\max_{V \in \mathbb{R}_*^{n \times r}} \mathcal{I}(W, Q) = \frac{1}{2} \sum_{i=1}^{r} \log \lambda_i,$$

where $(\lambda_i)_{i=1}^r$ *are the* r *dominant eigenvalues of the problem*

$$C_Y v = \lambda C_E v, \quad \lambda \in \mathbb{R}, \; v \in \mathbb{R}^n.$$

A solution to the optimization problem is given by the matrix V *with columns being eigenvectors* $(v_i)_{i=1}^r$ *associated to the eigenvalues* $(\lambda_i)_{i=1}^r$*. (Error estimator)*

Equivalences

The mutual information maximization is equivalent to:

- the maximization of the expected information gain
$$\max_{V \in \mathbb{R}_*^{n \times r}} \mathbb{E}_W \left(\mathrm{D}_{\mathrm{KL}} \left(P(Q|W) \,\|\, P(Q) \right) \right)$$
- the minimization of the entropy of the posterior distribution
$$\min_{V \in \mathbb{R}_*^{n \times r}} \mathrm{H} \left(P(Q|W = V^T y) \right)$$

Inference problem

Synthetic data

For $(t_i)_{i=1}^n$, $n = 500$, a uniformly drawn sample in $(-1, 1)$,

$$Y_{\text{ref}}(t_i) = A_{\text{ref}}(t_i) + E(t_i), \quad \forall i \in \{1, \dots, n\},$$

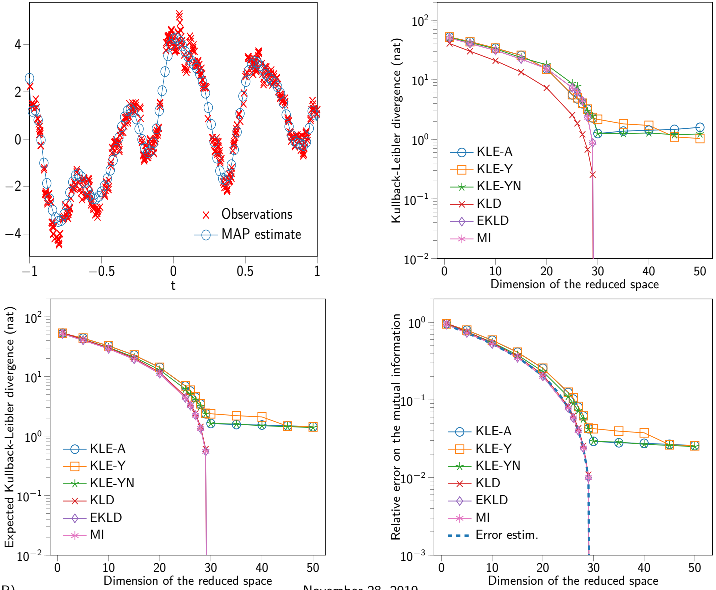
with $A_{\text{ref}} \sim \mathcal{N}(m_{\text{ref}}, C_{\text{ref}})$ and $E \sim \mathcal{N}(m_E, C_E)$.

Model

$$Y_i = \sum_{j=0}^{\mathrm{Nq}-1} T_j(t_i) Q_j + E(t_i), \quad \forall i \in \{1, \dots, n\},$$

with T_j the Chebyshev polynomial of order j and $\mathrm{Nq} = 30$.

Functionals versus the dimension of the reduced space



Inference problem: nonlinear models

Synthetic data

Given two random samples $(s_i)_{i=1}^n$ and $(t_i)_{i=1}^n$ being independent and uniformly distributed in $(-1, 1)$, with $n = 2000$,

$$Y_{\text{ref}}(s_i, t_i) = \exp(F_{\text{ref}}(s_i, t_i)) + E(s_i, t_i), \quad \forall i \in \{1, \dots, n\},$$

where $F_{\text{ref}} \sim \mathcal{N}(0, C_{\text{ref}})$, $E \sim \mathcal{N}(0, C_E)$.

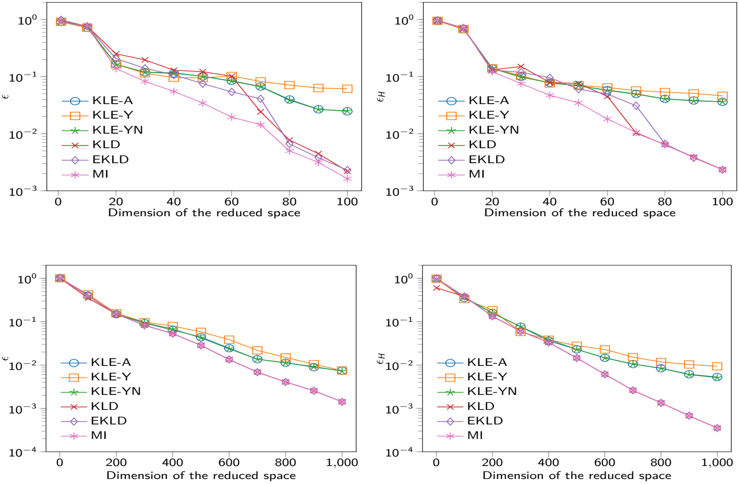
Model

$$Y_i = A_i(Q) + E(s_i, t_i), \quad \forall i \in \{1, \dots, n\},$$

where $A_i(Q) = \exp((BQ)_i)$, $Q \sim \mathcal{N}(0, C_Q)$, and $q = 30$.

- Columns of B: dominant eigenvectors of C_{ref}
- $C_Q = \text{diag}(\lambda_1, \dots, \lambda_q)$: dominant eigenvalues of C_{ref}

Errors versus the dimension of the reduced space $\sigma_{F_{\text{ref}}} = 0.301$ (top), $\sigma_{F_{\text{ref}}} = 1.501$ (bottom)



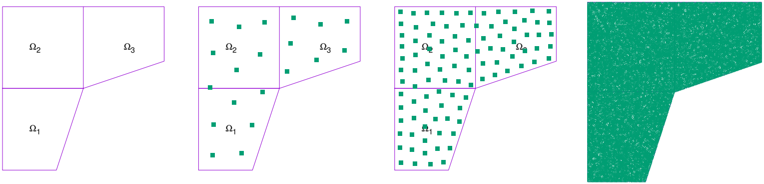
L_2 error on MAP point (left) and Frobenius error on Hessian at MAP.

Inference of conductivities

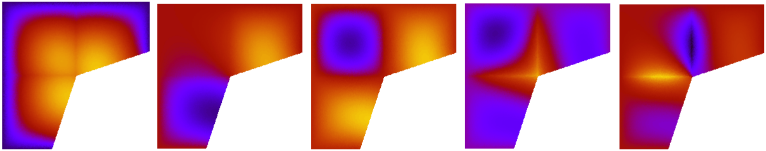
The model:

$$\nabla\left(\kappa(\boldsymbol{x})\nabla U(\boldsymbol{x})\right) = -1, \quad \kappa(\boldsymbol{x} \in \Omega_i) = \kappa_i,$$

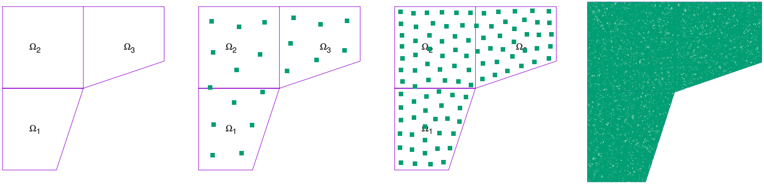
where $\log \kappa_i \sim N(0,1)$. Observed at $n = 32,000$ points with Gaussian noise.



Dominant modes of the projection:



Inference of conductivities



Convergence to unreduced MAP and Hessian:

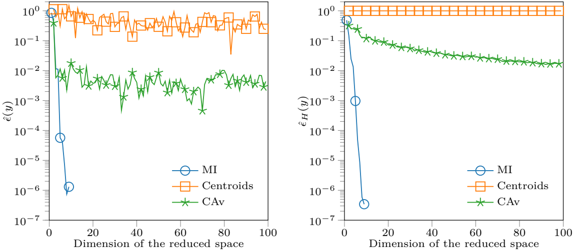


Figure 15: Convergence with the reduction dimension of the MI, Centroids and Cluster Averages errors on MAP ($\hat{\epsilon}(y)$, left) and Hessian ($\hat{\epsilon}_H(y)$, right). Case of high noise level $\sigma_\epsilon = 0.5$.

[Giraldi, OLM, Hoteit and Knio. Comp. Stat. & Data Anal., 2018.]

Table of contents

Conclusions and outlook

Summary

- Reduction approaches are instrumental in UQ and inference
- May concern both the model and the observations
- Reduction strategies should be goal-oriented
- Information theoretic reduction approaches are promising

Outlooks

- Selection of **observation features** for Bayesian inference
- **Goal-oriented** design of model reduction and experiments

Thank you