DE LA RECHERCHE À L'INDUSTRIE

www.cea.fr

Data-driven kernel representations for sampling with an unknown block dependence

G. Perrin[(1)],

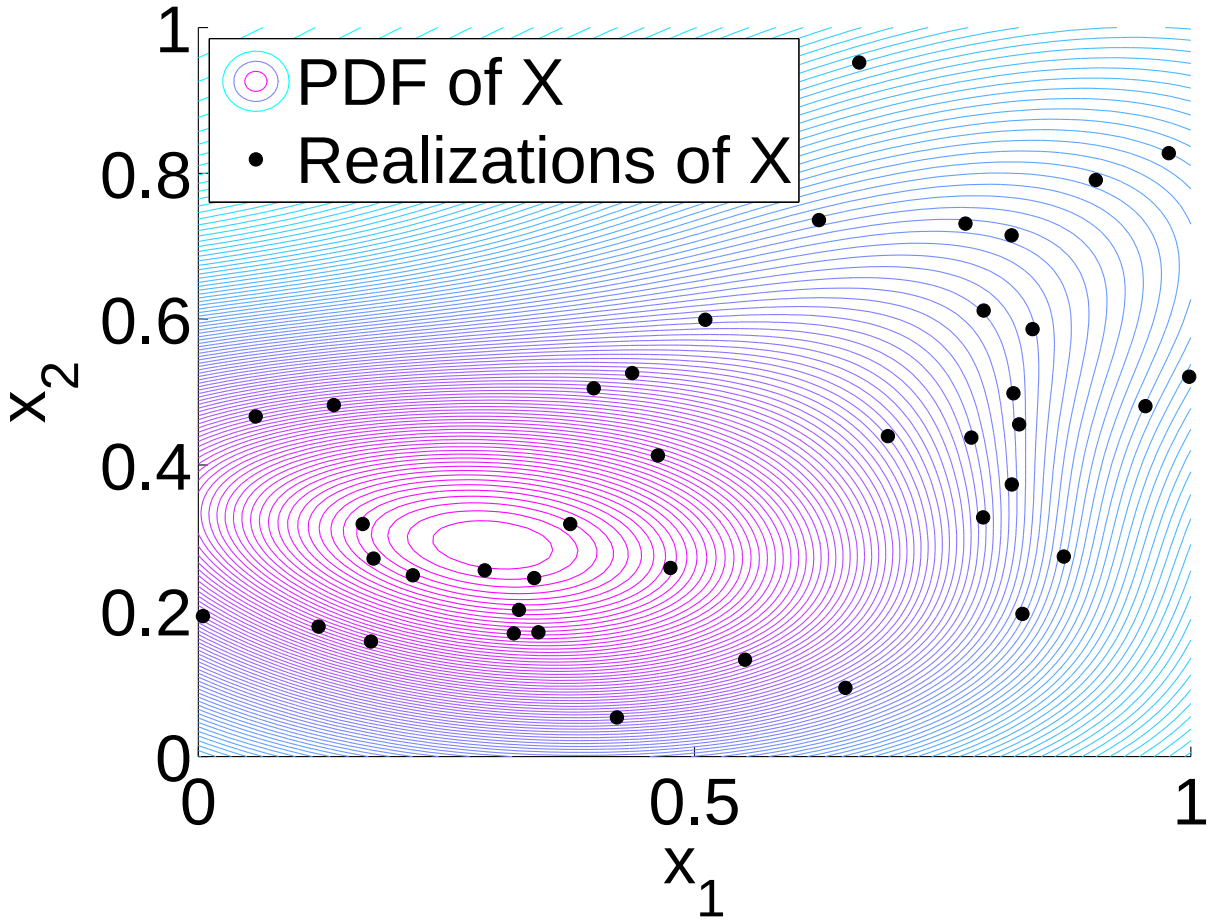
in collaboration with C. Soize[(2)] and N. Ouhbi[(3)]

[(1)] CEA/DAM/DIF, Arpajon, France
[(2)] Université Paris-Est, Marne-la-Vallée, France
[(3)] Innovation and Research Department, SNCF, Paris, France

MascotNum 2018 |

March, 22th 2018

Introduction



- When the maximal information about a d-dimensional random vector $\boldsymbol{X}$ is a set of N iid realizations, the **kernel density estimation** (KDE) is a widely used technique to infer the PDF of $\boldsymbol{X}$.

Introduction

- For instance, it can be used
  - to process the samples provided by a MCMC approach in the Bayesian calibration of a computational code [Berliner, 2001, Kaipio and Somersalo, 2004],
  - to approximate goal-oriented Sobol indices [Perrin and Defaux, 2018],
  - to optimize under uncertainties a particular code output [Soize and Ghanem, 2017].
- In practice, this technique is limited to cases when d is **small** (less than five in general).

Problematic

What could we propose to extend the validity of this technique to higher values of d ($d \sim 10 - 100$) with limited information ($N \sim 10d$ for instance)?

Outline

Theoretical framework

- Let $\boldsymbol{X} := \{\boldsymbol{X}(\omega),\ \omega \in \Omega\}$ be a second-order random vector defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with values in $\mathbb{R}^d$.
- The PDF of $\boldsymbol{X}$ is denoted by $p_{\boldsymbol{X}}$.
- The maximal available information about $p_{\boldsymbol{X}}$ is a set of $N > d$ independent and distinct realizations of $\boldsymbol{X}$, which are gathered in the deterministic set $\mathcal{S}(N) := \{\boldsymbol{X}(\omega_n),\ 1 \leq n \leq N\}$.
- Given these realizations of $\boldsymbol{X}$, the kernel estimator of $p_{\boldsymbol{X}}$ is

$$\widehat{p}_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{H}, \mathcal{S}(N)) = \frac{\det(\boldsymbol{H})^{-1/2}}{N} \sum_{n=1}^{N} K\left(\boldsymbol{H}^{-1/2}\left(\boldsymbol{x} - \boldsymbol{X}(\omega_n)\right)\right),$$

where $\det(\cdot)$ is the determinant operator, K is any function of $\mathcal{M}_1(\mathbb{R}^d, \mathbb{R}^+)$, and $\boldsymbol{H}$ is a $(d \times d)$-dimensional positive definite symmetric "**bandwidth matrix**".

Theoretical framework

- In this work, we focus on the classical case when K is the Gaussian multidimensional density ($\leftrightarrow$ "G-KDE"):

$$\widehat{p}_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{H}, \mathcal{S}(N)) = \frac{1}{N} \sum_{n=1}^{N} \phi\left(\boldsymbol{x}; \boldsymbol{X}(\omega_n), \boldsymbol{H}\right), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

- Here, $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{C})$ is the PDF of any $\mathbb{R}^d$-dimensional Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{C}$:

$$\phi\left(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{C}\right) := \frac{\exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T \boldsymbol{C}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right)}{(2\pi)^{d/2}\sqrt{\det(\boldsymbol{C})}}, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

$\Rightarrow p_{\boldsymbol{X}}$ is approximated by a mixture of N Gaussian PDFs, whose means are the available realizations of $\boldsymbol{X}$ and whose covariance matrices are all equal to $\boldsymbol{H}$

Identification of the kernel parameters

- By construction, $\boldsymbol{H}$ characterizes the **local contribution** of each realization of $\boldsymbol{X}$.
- Its value has to be optimized to minimize the difference between $p_{\boldsymbol{X}}$, which is unknown, and $\widehat{p}_{\boldsymbol{X}}(\cdot; \boldsymbol{H}, \mathcal{S}(N))$.
- The mean integrated squared error (MISE) performance criterion

$$\mathrm{MISE}(\boldsymbol{H}; d, N) = \mathbb{E}\left[\int_{\mathbb{R}^d} \left(p_{\boldsymbol{X}}(\boldsymbol{x}) - \widehat{p}_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{H}, \mathcal{S}(N))\right)^2 d\boldsymbol{x}\right]$$

  is generally considered to quantify such a difference.
- Given sufficient regularity conditions on $p_{\boldsymbol{X}}$, an asymptotic approximation of this criterion can be derived, leading to the commonly-used **Silverman bandwidth matrix** [Silverman, 1986]

$$\boldsymbol{H}^{\mathrm{Silv}}(d, N) := (h^{\mathrm{Silv}}(d, N))^2 \widehat{\boldsymbol{R}}_{\boldsymbol{X}}, \quad h^{\mathrm{Silv}}(d, N) := \left(\frac{1}{N}\frac{4}{(d+2)}\right)^{\frac{1}{d+4}},$$

with $\widehat{\boldsymbol{R}}_{\boldsymbol{X}}$ the empirical estimation of the covariance matrix of $\boldsymbol{X}$.

Identification of the kernel parameters

- In practice, it is generally observed that, for fixed values of N, $\boldsymbol{H}^{\mathsf{Silv}}(d,N)$ **overestimates** the scattering of $p_{\boldsymbol{X}}$.
- To circumvent this problem, the LOO expression of the likelihood,

$$\mathcal{L}^{\mathsf{LOO}}(\mathcal{S}(N)|\boldsymbol{H}) := \prod_{n=1}^{N} \frac{1}{N-1} \sum_{m=1,m\neq n}^{N} \phi_{n,m}(\boldsymbol{H}),$$

$$\phi_{n,m}(\boldsymbol{H}) := \phi\left(\boldsymbol{X}(\omega_n); \boldsymbol{X}(\omega_m), \boldsymbol{H}\right), \quad 1 \leq n, m \leq N,$$

  can instead directly be used to identify $\boldsymbol{H}$ [van der Laan et al., 2004].
- In this presentation, we will focus on the **maximum likelihood estimate** of $\boldsymbol{H}$, denoted by

$$\boldsymbol{H}^{\mathsf{MLE}}(d,N) := \arg \max_{\boldsymbol{H}\in\mathbb{M}^{+}(d)} \mathcal{L}^{\mathsf{LOO}}(\mathcal{S}(N)|\boldsymbol{H}).$$

Imposing correlation constraints

- Considering that the best available approximations of the true mean and covariance matrix of $\boldsymbol{X}$ are given by their empirical estimations, the former expression can be slightly modified.
- Indeed, if the PDF of $\boldsymbol{X}$ is equal to

$$\widetilde{p}_{\boldsymbol{X}}(\cdot; \boldsymbol{H}, \mathcal{S}(N)) := \frac{1}{N} \sum_{n=1}^{N} \phi\left(\cdot; \boldsymbol{A}\boldsymbol{X}(\omega_n) + \boldsymbol{\beta}, \boldsymbol{H}\right),$$

$$\boldsymbol{\beta} := (\boldsymbol{I}_d - \boldsymbol{A})\widehat{\boldsymbol{\mu}}, \quad \boldsymbol{H} := \widehat{\boldsymbol{R}}_{\boldsymbol{X}} - \frac{N-1}{N} \boldsymbol{A}\widehat{\boldsymbol{R}}_{\boldsymbol{X}}\boldsymbol{A}^T,$$

  the mean and the covariance matrix of $\boldsymbol{X}$ are equal to $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{R}}_{\boldsymbol{X}}$ respectively [Perrin et al., 2018].
- Given $\mathcal{S}(N)$, the G-KDE of the PDF of $\boldsymbol{X}$ under constraints on its mean and its covariance matrix will be denoted by $\widetilde{p}_{\boldsymbol{X}}(\cdot; \boldsymbol{H}^{\mathsf{MLE}}(d, N), \mathcal{S}(N))$ in the following.

Difficulties in high dimension

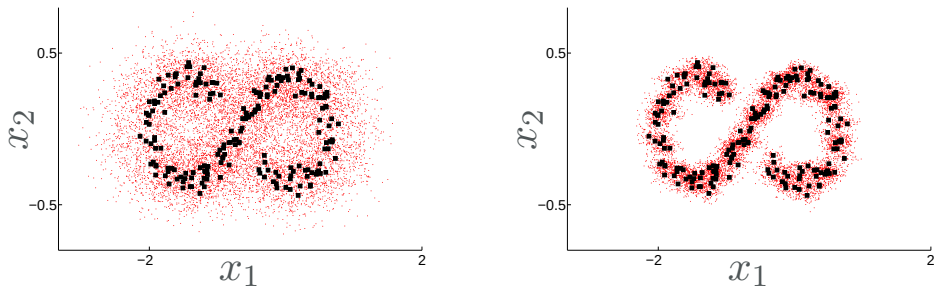
- In practice, when considering the nonparametric modelling of high dimensional random vectors ($d \sim 10 - 100$) with limited information ($N \sim 10d$ for instance), we observe that $\boldsymbol{H}^{\mathsf{MLE}}(d, N)$ is very close to $\widehat{\boldsymbol{R}}_{\boldsymbol{X}}$.
- This means that we are approximating the PDF of $\boldsymbol{X}$ as a **unique** Gaussian PDF, whose parameters correspond to the empirical mean and covariance matrix of $\boldsymbol{X}$:

$$\lim_{\boldsymbol{H} \to \widehat{\boldsymbol{R}}_{\boldsymbol{X}}} \widetilde{p}_{\boldsymbol{X}}(\cdot; \boldsymbol{H}, \mathcal{S}(N)) = \phi(\cdot; \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{R}}_{\boldsymbol{X}}).$$

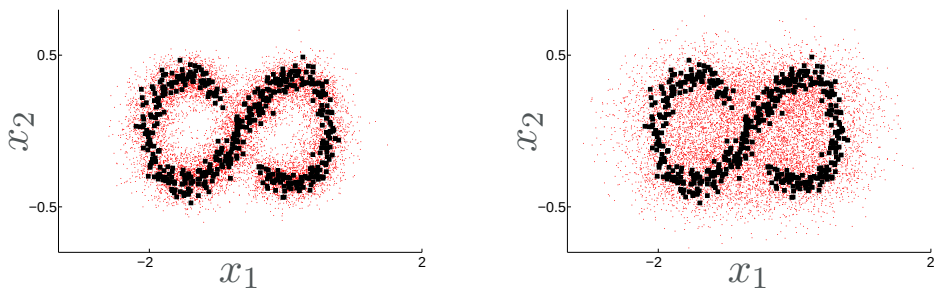
- This could prevent us from recovering the subset of $\mathbb{R}^d$ on which $\boldsymbol{X}$ is actually concentrated.

Difficulties in high dimension - the Lemniscate function

$\boldsymbol{X}^{(2D)} = (X_1^{\mathsf{L}} + 0.05\xi_1, X_2^{\mathsf{L}} + 0.05\xi_2, \xi_3, \ldots, \xi_d)$,
$\boldsymbol{X}^{\mathsf{L}} \leftrightarrow$ Lemniscate function, ξ_i are iid standard Gaussian r.v.



(a) $d = 2$, $\boldsymbol{H}^{\mathsf{Silv}}$ (b) $d = 2$, $\boldsymbol{H}^{\mathsf{MLE}}$

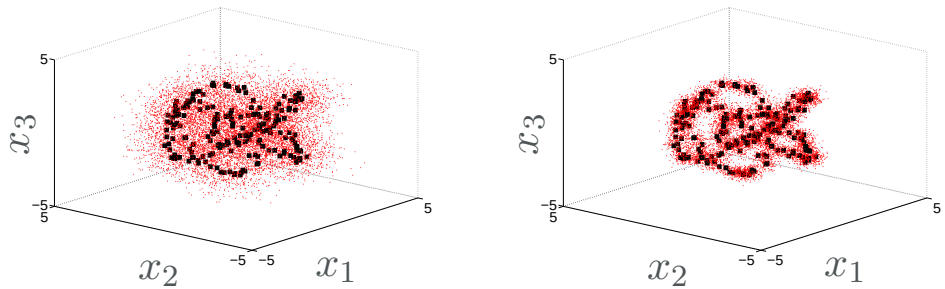


(c) $d = 3$, $\boldsymbol{H}^{\mathsf{MLE}}$ (d) $d = 5$, $\boldsymbol{H}^{\mathsf{MLE}}$

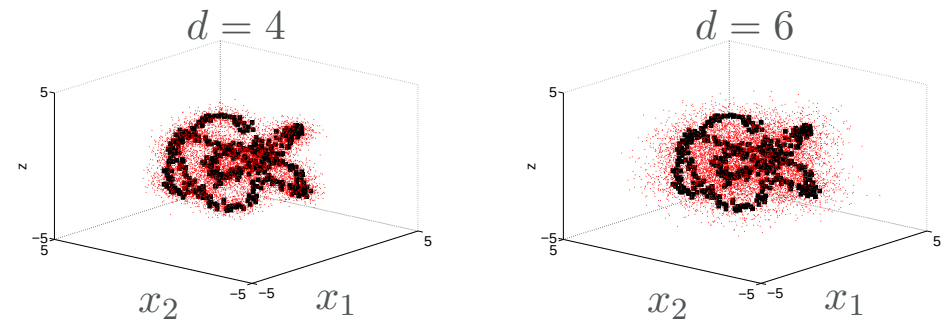
Figure: $N = 500$ given values of $\boldsymbol{X}^{(2D)}$ (big black squares) and 10^4 additional values (small red points) generated from a G-KDE approach for several values of d.

Difficulties in high dimension - the four branches clover-knot function

$\boldsymbol{X}^{(3D)} = (X_1^{\mathsf{FB}} + \xi_1, \ldots, X_3^{\mathsf{FB}} + \xi_3, \xi_4, \ldots, \xi_d)$,
$\boldsymbol{X}^{\mathsf{FB}} \leftrightarrow$ Four branch clover-knot function, ξ_i iid standard Gaussian r.v.



(a) $d = 3$, $\boldsymbol{H}^{\mathsf{Silv}}$ (b) $d = 3$, $\boldsymbol{H}^{\mathsf{MLE}}$



(c) $d = 4$, $\boldsymbol{H}^{\mathsf{MLE}}$ (d) $d = 6$, $\boldsymbol{H}^{\mathsf{MLE}}$

Figure: $N = 500$ given values of $\boldsymbol{X}^{(3D)}$ (big black squares) and 10^4 additional values (small red points) generated from a G-KDE approach for several values of d.

Outline

Block-by-block decomposition

- The idea is to identify groups of components of $\boldsymbol{X}$ that can **reasonably** be considered as statistically **independent**, if they exist.
- Given a decomposition of $\boldsymbol{X}$ in N_b blocks $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N_b)}$, PDF $p_{\boldsymbol{X}}$ is approximated as the product of the nonparametric estimations of the PDFs associated with each sub-vector of $\boldsymbol{X}$:

$$p_{\boldsymbol{X}} \approx \prod_{\ell=1}^{N_b} \widetilde{p}_{\boldsymbol{X}^{(\ell)}}(\cdot; \boldsymbol{H}^{(\ell)}, \mathcal{S}(N)).$$

- Instead of using statistical tests, we propose to search these groups by looking for the minimum of a cross-validation AIC that is associated with each block formation [Perrin et al., 2018].

Hypotheses

1. $\boldsymbol{X}$ is **centred** and **uncorrelated**: $\widehat{\boldsymbol{\mu}}_{\boldsymbol{X}} = \boldsymbol{0}, \quad \widehat{\boldsymbol{R}}_{\boldsymbol{X}} = \boldsymbol{I}_d.$
2. $\boldsymbol{H}_\ell$ is parametrized by a **unique scalar**: $\boldsymbol{H}_\ell = h_\ell^2 \boldsymbol{I}_{d_\ell}, \, 0 < h_\ell \leq 1.$

Identification of the block-by-block decomposition

- For any $\boldsymbol{b}$ in $\{1, \ldots, d\}^d$, b_i can be used as a block index for the i^{th} component of $\boldsymbol{X}$.
- This means that if $b_i = b_j$, X_i and X_j are supposed to be **dependent** and belong to the **same** block. On the contrary, if $b_i \neq b_j$, X_i and X_j are supposed to be **independent** and they can belong to two **different** blocks.
- There exists a bijection between the set of all block by block decompositions of $\boldsymbol{X}$ and the set

$$\mathbb{B}(d) := \left\{ \begin{array}{c} \boldsymbol{b} \in \{1, \ldots, d\}^d \mid b_1 = 1, \\ 1 \leq b_j \leq 1 + \max_{1 \leq i \leq j-1} b_i, \ 2 \leq j \leq d \end{array} \right\}.$$

Difficulty

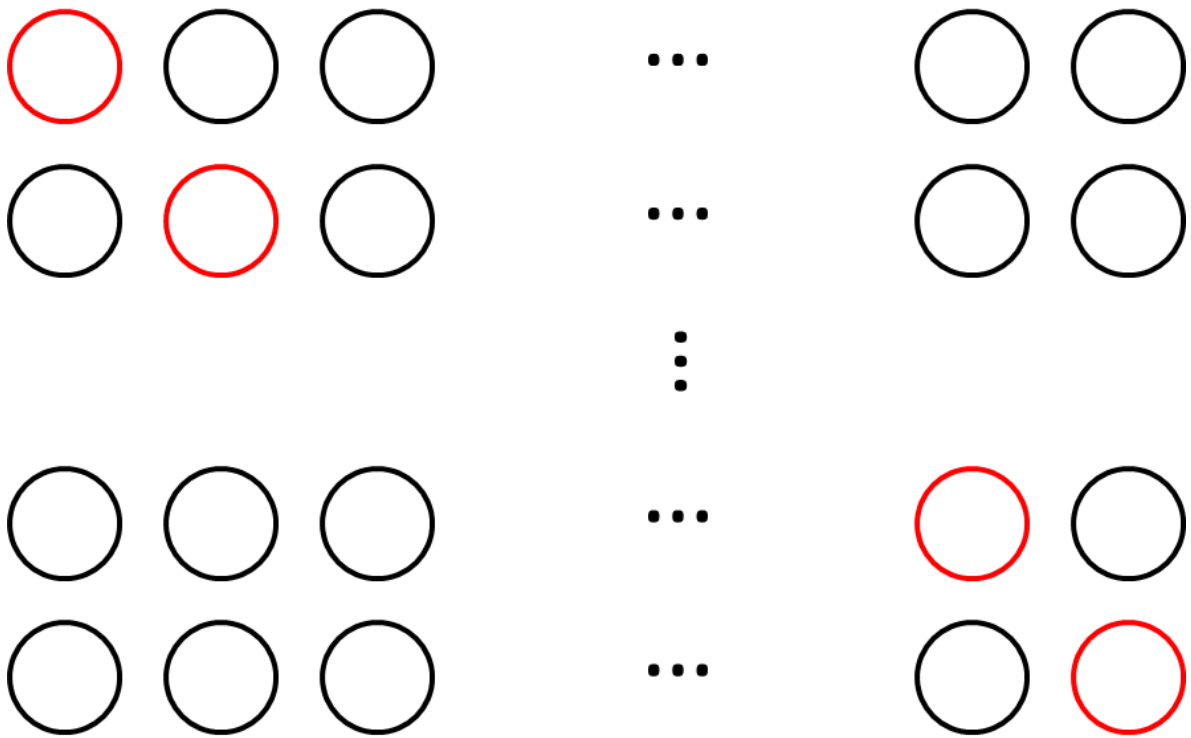
The cardinal of $\mathbb{B}(d)$ increases exponentially with d.

Identification of the block-by-block decomposition

Greedy identification - initialization

$$\boldsymbol{b} = (1, \ldots, 1)$$

Greedy identification - first loop



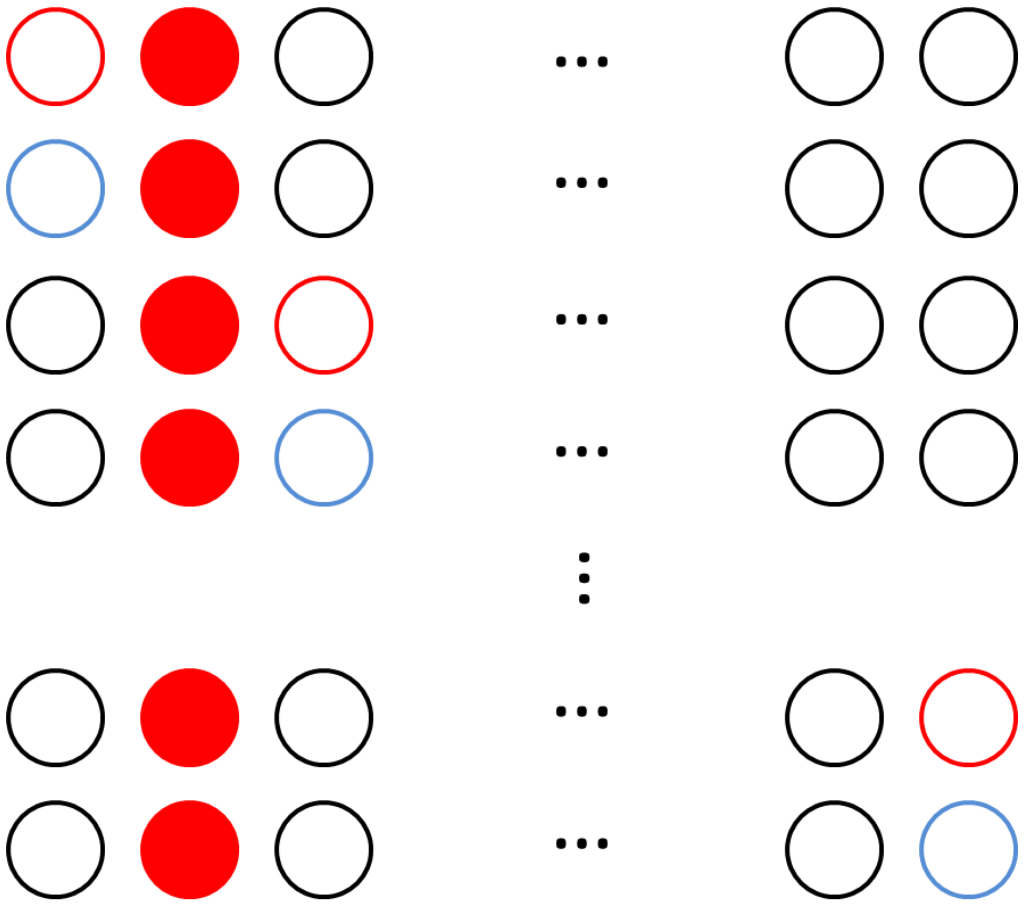
$$\boldsymbol{b} \in \{(b_1, \ldots, b_d) \mid b_i = 2,\ b_{j\neq i} = 1,\ 1 \leq i \leq d\}$$

Identification of the block-by-block decomposition

Greedy identification - first fixed point

$$\boldsymbol{b} = (1, 2, 1, \ldots, 1)$$

Greedy identification - second loop



$$\boldsymbol{b} \in \{(b_1, \ldots, b_d) \mid b_2 = 2,\ b_i \in \{2, 3\},\ b_{j \neq i, j \neq 2} = 1,\ i \neq 2\}$$

Identification of the block-by-block decomposition

Greedy identification - second fixed point

$$\boldsymbol{b} = (1, 2, 1, \ldots, 1, 3, 1)$$

Identification of the block-by-block decomposition

Greedy identification - last fixed point

$$\boldsymbol{b} = (1, 2, 3, \ldots, 1, 3)$$

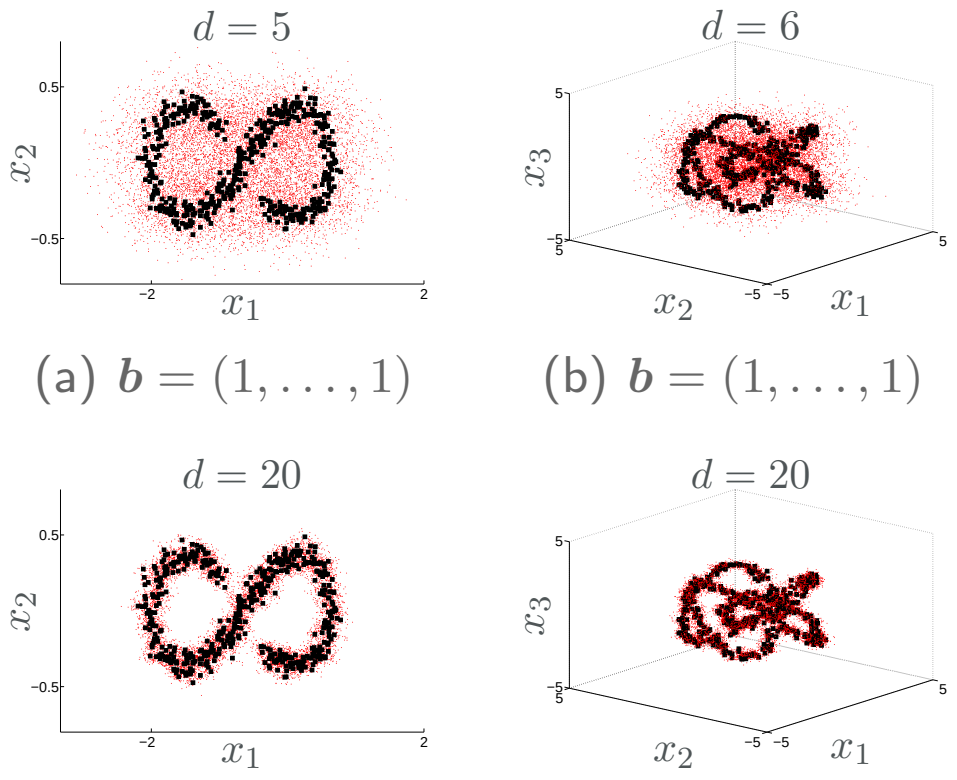
Interest of the proposed decomposition in high dimension

- The greedy algorithm allows the identification of good values for $\boldsymbol{b}$ at a reduced computational cost.
- We verified on a series of test cases that the algorithm was able to identify the correct block decomposition with $2 \leq d \leq 100$ and $N = 10d$ in a reasonable computational time.
- Evolutionary algorithms can also be used to address problems in higher dimensions.

d	1	2	3	4	5	6	7	8	9	10
$\sharp\mathbb{B}(d)$	1	2	5	15	52	203	877	4140	21147	115975
$N^{\max}_{\text{greedy}}(d)$	1	3	8	17	31	51	78	113	157	211

Table: Evolution of the number of elements of $\mathbb{B}(d)$, $\sharp\mathbb{B}(d)$, and the maximum number of configurations tested by the greedy algorithm, $N^{\max}_{\text{greedy}}(d)$, with respect to d.

Interest of the proposed decomposition in high dimension



(a) $\boldsymbol{b} = (1, \ldots, 1)$ (b) $\boldsymbol{b} = (1, \ldots, 1)$

(c) $\boldsymbol{b} = (1, 1, 2, 3, \ldots, 18)$ (d) $\boldsymbol{b} = (1, 1, 1, 2, 3, \ldots, 17)$

Figure: $N = 500$ given values of $\boldsymbol{X}^{(3D)}$ (big black squares) and 10^4 additional values (small red points) generated from a G-KDE approach for several values of d.

Outline

Physical description of the problem

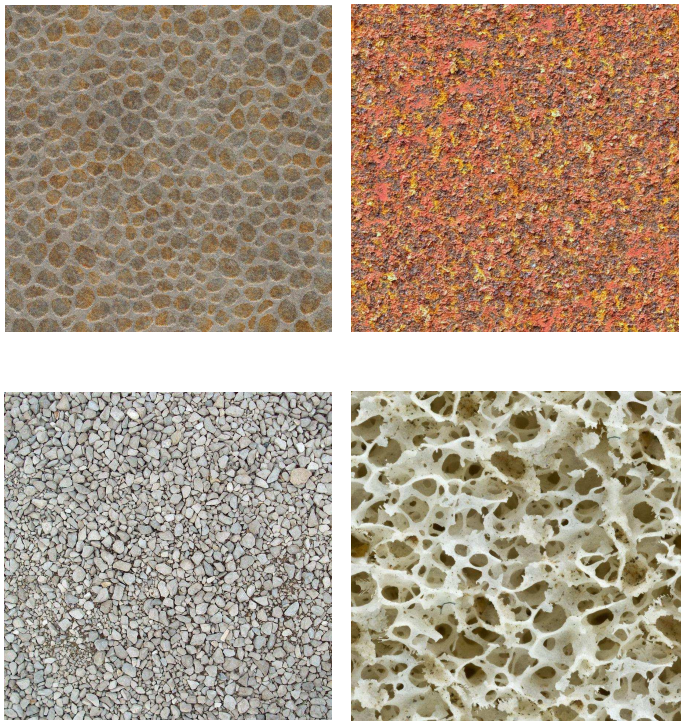


Figure: Four heterogeneous media

Problematic

How to infer the statistical properties of a random medium from a limited number of indirect measurements?

Physical description of the problem

- We are interested in the identification of the mechanical properties of a **heterogeneous elastic medium**.
- Several experimental tests are performed on a series of M specimens made of the same material. Let $\mathcal{V}$ be their volume.
- For each experiment, the applied force field is supposed to be **imposed**, and the induced displacement field is **measured on the contours** of the specimens only.
- In parallel, for given properties of the considered medium, it is possible to approximate (using the Finite Element Method) the displacements that are induced by the experimental force field.
- In this work, we focus on the estimation of 5 quantities gathered in the vector $\boldsymbol{z} = (\lambda, \ell_1, \ell_2, \mu_\nu, \mu_E)$, where λ is a fluctuation level, ℓ_1 and ℓ_2 are two correlation lengths, and μ_ν and μ_E are the means of the Poisson coefficient and the Young modulus respectively.

Mathematical description of the problem

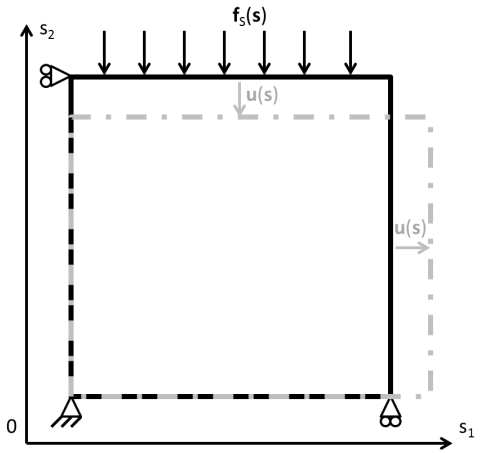
- Let $\boldsymbol{X}$ be the elasticity field characterizing the mechanical properties of the material that constitutes the specimens.
- $\boldsymbol{X}$ is supposed to be **random**, and we assume that it belongs to a known class of parametric random fields, such that:

$$\boldsymbol{X} = \{\boldsymbol{X}(\boldsymbol{s}, \omega; \boldsymbol{z}^\star),\ \boldsymbol{s} \in \mathcal{V},\ \omega \in \Omega\},$$

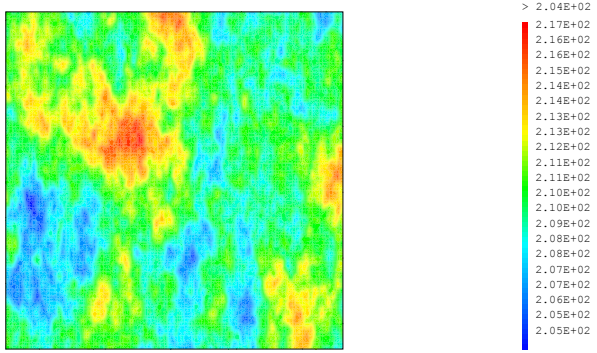
  where $\boldsymbol{z}^\star \in \mathbb{Z}$ is unknown.
- $\boldsymbol{X}$ is not a real-valued random field, but a tensor-valued random field, and its different components cannot be identified separately due to algebraic constraints.
- Let $\boldsymbol{u}(\boldsymbol{X}(\omega; \boldsymbol{z}))$ be the induced displacement on the contour of the specimen associated with the particular realization $\boldsymbol{X}(\omega; \boldsymbol{z})$ of $\boldsymbol{X}(\boldsymbol{z})$.
- This displacement can be decomposed in two contributions :

$$\boldsymbol{u}(\boldsymbol{s}; \boldsymbol{X}(\omega; \boldsymbol{z})) = \widehat{\boldsymbol{u}}(\boldsymbol{s}; \mathbb{E}[\boldsymbol{X}(\boldsymbol{z})]) + \widetilde{\boldsymbol{u}}(\boldsymbol{s}; \boldsymbol{X}(\omega; \boldsymbol{z})).$$

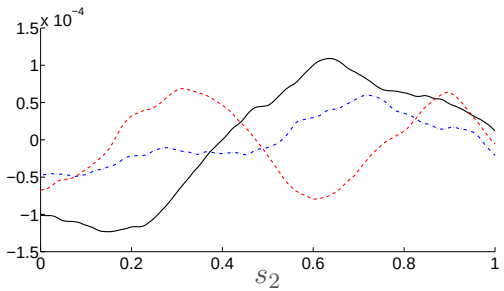
Illustration of the problem



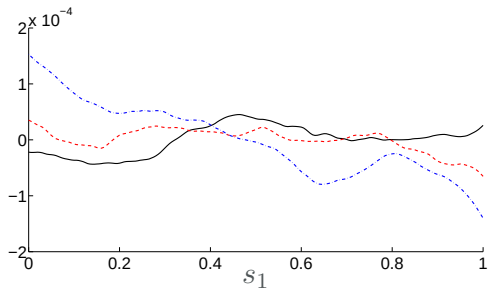
(a) Studied mechanical phenomenon



(b) One possible evolution of the Young modulus



(c) Three measured evolutions of $\widetilde{u}_1(\boldsymbol{z}^\star)$



(d) Three measured evolutions of $\widetilde{u}_2(\boldsymbol{z}^\star)$

Bayesian formulation

- $z^\star$ is modelled by a **random vector**, denoted by $\boldsymbol{Z}$, to take into account the fact that its values are unknown. Let $f_{\boldsymbol{Z}}$ be its PDF.
- Let $\boldsymbol{Y}(\boldsymbol{z})$ be a d_y-dimensional random vector that condenses the statistical properties of $\boldsymbol{u}(\boldsymbol{X}(\boldsymbol{z}))$ and $f_{\boldsymbol{Y}(\boldsymbol{z})}$ be its PDF.
- M independent realizations of $\boldsymbol{Y}(\boldsymbol{z}^\star)$ are gathered in the set $\mathbb{Y} := \{\boldsymbol{Y}(\omega_1; \boldsymbol{z}^\star), \ldots, \boldsymbol{Y}(\omega_M; \boldsymbol{z}^\star)\}$ (one for each specimen).
- Using the Bayes theorem, it comes:

$$f_{\boldsymbol{Z}|\mathbb{Y}}(\boldsymbol{z}) = \frac{\mathcal{L}_{\mathbb{Y}}(\boldsymbol{z}) f_{\boldsymbol{Z}}(\boldsymbol{z})}{\mathbb{E}\left[\mathcal{L}_{\mathbb{Y}}(\boldsymbol{Z})\right]}, \ \boldsymbol{z} \in \mathbb{R}^{d_z}.$$

There, $\mathcal{L}_{\mathbb{Y}}(\boldsymbol{z})$ is the **likelihood function**.

Approximation of the likelihood

$$\mathcal{L}_{\mathbb{Y}}(\boldsymbol{z}) = \prod_{m=1}^{M} f_{\boldsymbol{Y}(\boldsymbol{z})}(\boldsymbol{Y}(\omega_m; \boldsymbol{z}^\star)), \ \boldsymbol{z} \in \mathbb{R}^{d_z}.$$

Independent estimation

$$\mathcal{L}_{\mathbb{Y}}(\boldsymbol{z}) \approx \prod_{m=1}^{M} \widetilde{f}_{\boldsymbol{Y}(\boldsymbol{z})}(\boldsymbol{Y}(\omega_m; \boldsymbol{z}^\star)),$$

where $\widetilde{f}_{\boldsymbol{Y}(\boldsymbol{z})}$ is the kernel estimator of $f_{\boldsymbol{Y}(\boldsymbol{z})}$ based on N independent realizations of $\boldsymbol{Y}(\boldsymbol{Z})|\boldsymbol{Z} = \boldsymbol{z}$.

For each value of $\boldsymbol{z}$, N evaluations of the code are required to approximate $\mathcal{L}_{\mathbb{Y}}(\boldsymbol{z}) \Rightarrow$ as the likelihood function has to be evaluated a high number of times to get precise information about the PDF of $\boldsymbol{Z}|\mathbb{Y}$, this computational cost is generally not affordable.

Approximation of the likelihood

Joint estimation

$$\mathcal{L}_{\mathbb{Y}}(\boldsymbol{z}) \approx \prod_{m=1}^{M} \widetilde{f}_{\boldsymbol{Y}(\boldsymbol{z})}(\boldsymbol{Y}(\omega_m; \boldsymbol{z}^\star)),$$

$$\widetilde{f}_{\boldsymbol{Y}(\boldsymbol{z})}(\boldsymbol{y}) = \frac{\widetilde{f}_{\boldsymbol{Y},\boldsymbol{Z}}(\boldsymbol{y},\boldsymbol{z})}{\int_{\mathbb{R}^{d_y}} \widetilde{f}_{\boldsymbol{Y},\boldsymbol{Z}}(\boldsymbol{v},\boldsymbol{z})d\boldsymbol{v}}$$

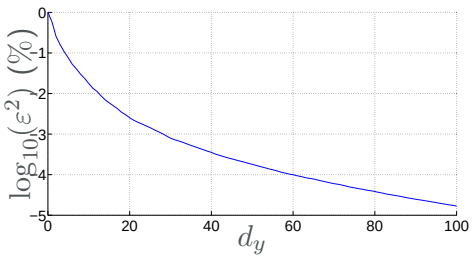
where $\widetilde{f}_{\boldsymbol{Y},\boldsymbol{Z}}$ is the kernel estimator of the PDF of $(\boldsymbol{Y}(\boldsymbol{Z}),\boldsymbol{Z})$ based on N independent realizations of $(\boldsymbol{Y}(\boldsymbol{Z}),\boldsymbol{Z})$.

$\Rightarrow$ only N evaluations of the code are needed to approximate the whole function $\mathcal{L}_{\mathbb{Y}}$.

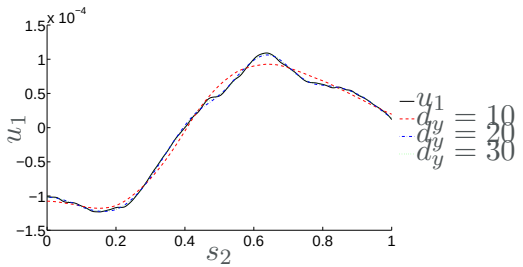
Remark: using Gaussian kernels, the expression of $\widetilde{f}_{\boldsymbol{Y}(\boldsymbol{z})}$ is **explicit** once $\widetilde{f}_{\boldsymbol{Y},\boldsymbol{Z}}$ is known.

Influence of the block decomposition on the calibration results

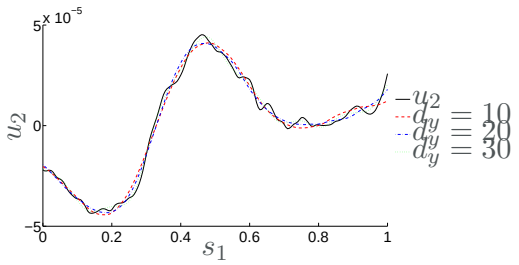
- $N = 1000$ code evaluations are carried out to infer the value of $\boldsymbol{z}^\star$.
- Coefficients μ_ν and μ_E are estimated using preliminary comparisons to the homogeneous case.
- The components of $\boldsymbol{Y}(\boldsymbol{z})$ correspond to the d_y first components of the KL decomposition of $\boldsymbol{u}(\boldsymbol{X}(\boldsymbol{z}))$.



(e) Error convergence



(f) Reduced basis representation of $u_1(s_2)$

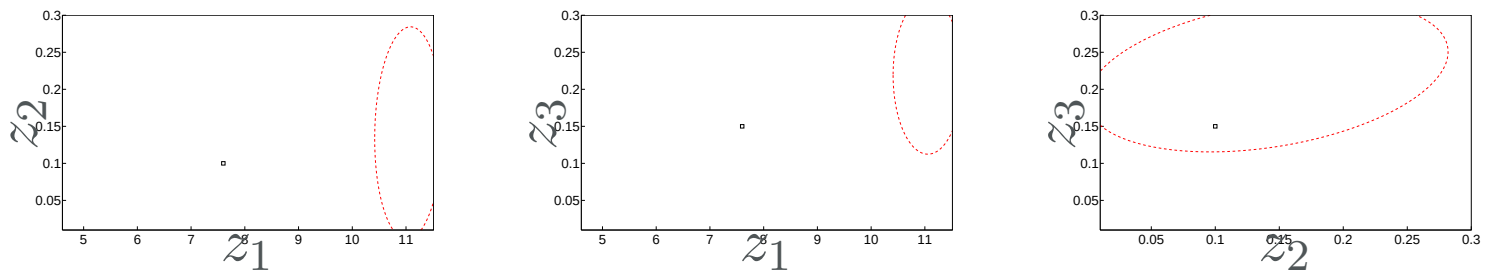


(g) Reduced basis representation of $u_2(s_1)$

Figure: Evolution of the projection error with respect to d_y.

Influence of the block decomposition on the calibration results

- $N = 1000$ code evaluations are carried out to infer the value of $\boldsymbol{z}^\star$.
- Coefficients μ_ν and μ_E are estimated using preliminary comparisons to the homogeneous case.
- The components of $\boldsymbol{Y}(\boldsymbol{z})$ correspond to the d_y first components of the KL decomposition of $\boldsymbol{u}(\boldsymbol{X}(\boldsymbol{z}))$.

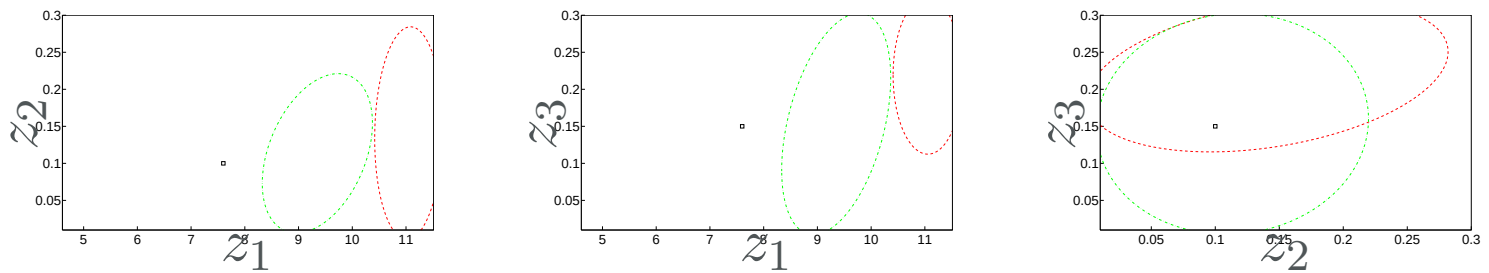


(a) $(-\log(\lambda), \ell_1)$ (b) $(-\log(\lambda), \ell_2)$ (c) (ℓ_1, ℓ_2)

Figure: White square: reference value. Red: 95% credible ellipses using $d_y = 23$ components (corresponding to a projection error of 0.1%).

Influence of the block decomposition on the calibration results

- $N = 1000$ code evaluations are carried out to infer the value of $\boldsymbol{z}^\star$.
- Coefficients μ_ν and μ_E are estimated using preliminary comparisons to the homogeneous case.
- The components of $\boldsymbol{Y}(\boldsymbol{z})$ correspond to the d_y first components of the KL decomposition of $\boldsymbol{u}(\boldsymbol{X}(\boldsymbol{z}))$.

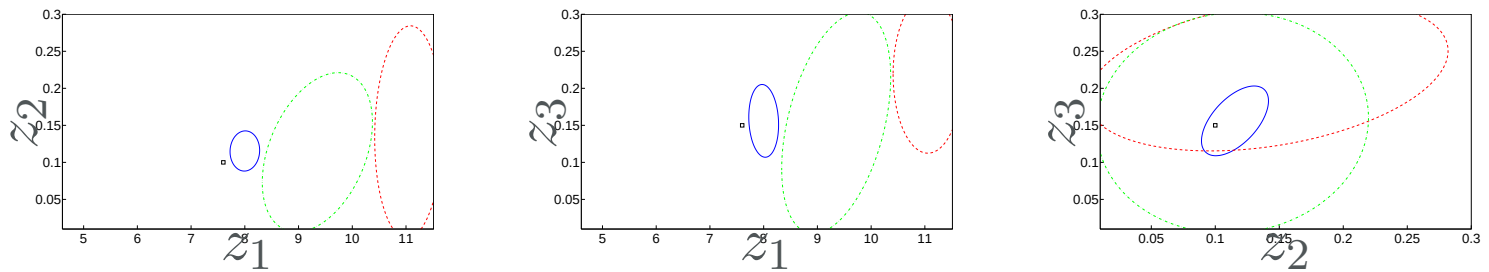


(a) $(-\log(\lambda), \ell_1)$ (b) $(-\log(\lambda), \ell_2)$ (c) (ℓ_1, ℓ_2)

Figure: White square: reference value. Red: 95% credible ellipses using $d_y = 23$ components (corresponding to a projection error of 0.1%). Green: 95% credible ellipses using only $d_y = 5$ components.

Influence of the block decomposition on the calibration results

- $N = 1000$ code evaluations are carried out to infer the value of $\boldsymbol{z}^\star$.
- Coefficients μ_ν and μ_E are estimated using preliminary comparisons to the homogeneous case.
- The components of $\boldsymbol{Y}(\boldsymbol{z})$ correspond to the d_y first components of the KL decomposition of $\boldsymbol{u}(\boldsymbol{X}(\boldsymbol{z}))$.



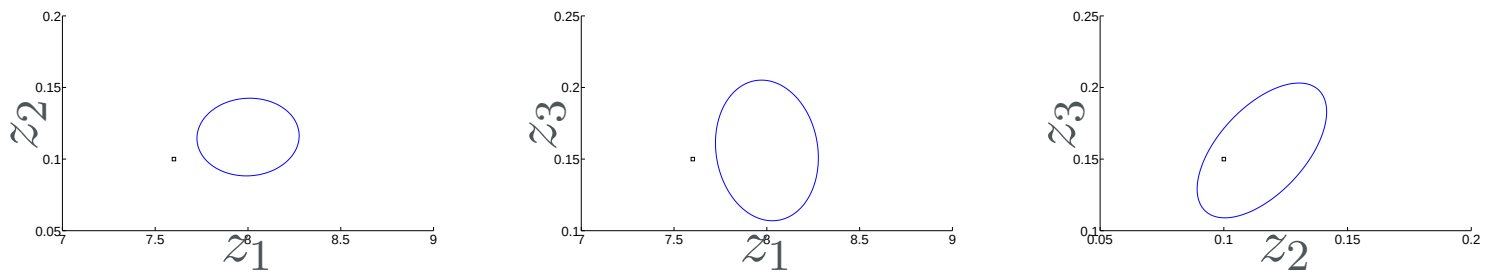
(a) $(-\log(\lambda), \ell_1)$ (b) $(-\log(\lambda), \ell_2)$ (c) (ℓ_1, ℓ_2)

Figure: White square: reference value. Red: 95% credible ellipses using $d_y = 23$ components (corresponding to a projection error of 0.1%). Green: 95% credible ellipses using only $d_y = 5$ components. Blue: 95% credible ellipses using $d_y = 23$ components, with an optimization of the block decomposition of $\boldsymbol{Y}(\boldsymbol{Z})|\boldsymbol{Z}$.

Adaptive calibration

Adaptive calibration

- Step 1: $N = 1000$ code evaluations are carried out.

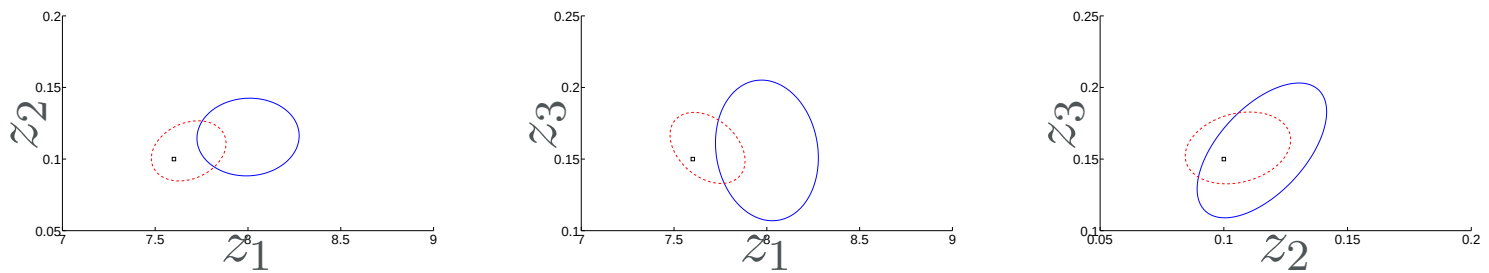


(a) $(-\log(\lambda), \ell_1)$ (b) $(-\log(\lambda), \ell_2)$ (c) (ℓ_1, ℓ_2)

Figure: White square: reference value. Blue: results of step 1, with $d_y = 23$ and 23 blocks for the PDF of $\boldsymbol{Y}(\boldsymbol{Z})|\boldsymbol{Z}$.

Adaptive calibration

- Step 1: $N = 1000$ code evaluations are carried out.
- Step 2: 889 new points in the likely region (provided by the calibration results of step 1) of $z^\star$ are added to the learning set.

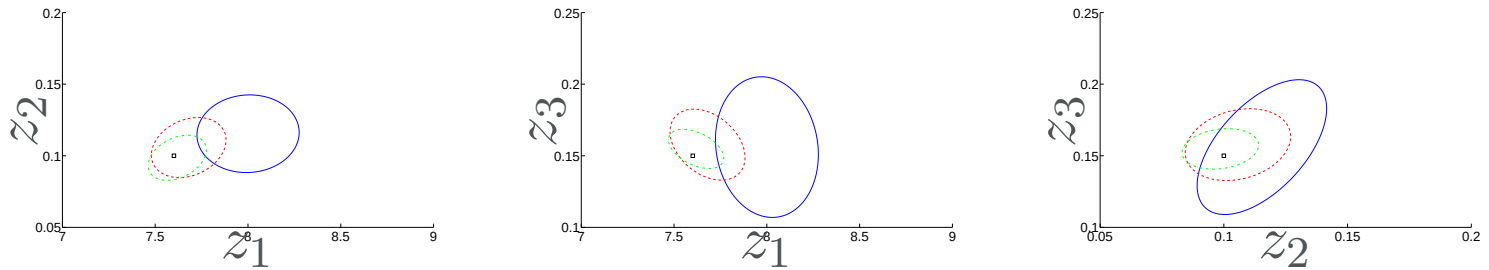


(a) $(-\log(\lambda), \ell_1)$ (b) $(-\log(\lambda), \ell_2)$ (c) (ℓ_1, ℓ_2)

Figure: White square: reference value. Blue: results of step 1, with $d_y = 23$ and 23 blocks for the PDF of $\boldsymbol{Y}(\boldsymbol{Z})|\boldsymbol{Z}$. Red: results of step 2, with $d_y = 23$ and 8 blocks for the PDF of $\boldsymbol{Y}(\boldsymbol{Z})|\boldsymbol{Z}$.

Adaptive calibration

- Step 1: $N = 1000$ code evaluations are carried out.
- Step 2: 889 new points in the likely region (provided by the calibration results of step 1) of $\boldsymbol{z}^\star$ are added to the learning set.
- Step 3: 631 new points in the likely region (provided by the calibration results of step 2) of $\boldsymbol{z}^\star$ are added to the learning set.



(a) $(-\log(\lambda), \ell_1)$ (b) $(-\log(\lambda), \ell_2)$ (c) (ℓ_1, ℓ_2)

Figure: White square: reference value. Blue: results of step 1, with $d_y = 23$ and 23 blocks for the PDF of $\boldsymbol{Y}(\boldsymbol{Z})|\boldsymbol{Z}$. Red: results of step 2, with $d_y = 23$ and 8 blocks for the PDF of $\boldsymbol{Y}(\boldsymbol{Z})|\boldsymbol{Z}$. Green: results of step 3, with $d_y = 23$ and 4 blocks for the PDF of $\boldsymbol{Y}(\boldsymbol{Z})|\boldsymbol{Z}$.

Outline

Conclusion

- This work considers the challenging problem of identifying complex PDFs when the maximal available information is a set of independent realizations.
- In that prospect, the multidimensional G-KDE method plays a key role, as it presents a good compromise between complexity and efficiency.
- Two adaptations of this method have been presented to deal with high dimensional random vector:
  - a modified formalism is presented to make the mean and the covariance matrix of the estimated PDF equal to their empirical estimations.
  - tensorized representations are proposed, which are based on the identification of a block by block dependence structure of the random vectors of interest.
- The interest of these two adaptations has been illustrated for the identification of the mechanical properties of a random medium.

Conclusion

- This work considers the challenging problem of identifying complex PDFs when the maximal available information is a set of independent realizations.
- In that prospect, the multidimensional G-KDE method plays a key role, as it presents a good compromise between complexity and efficiency.
- Two adaptations of this method have been presented to deal with high dimensional random vector:
  - a modified formalism is presented to make the mean and the covariance matrix of the estimated PDF equal to their empirical estimations.
  - tensorized representations are proposed, which are based on the identification of a block by block dependence structure of the random vectors of interest.
- The interest of these two adaptations has been illustrated for the identification of the mechanical properties of a random medium.

Thank you for your attention! Questions?

References I

Berliner, L. M. (2001).
Monte carlo based ensemble forecasting.
Statistics and Computing, 11.

Kaipio, J. P. . and Somersalo, E. (2004).
Statistics and Computational Inverse Problems.
Springer, New York.

Perrin, G. and Defaux, G. (2018).
Efficient evaluation of reliability-oriented sensitivity indices.
Submitted to Reliability Engineering and System Safety.

Perrin, G., Soize, C., and Ouhbi, N. (2018).
Data-driven kernel representations for sampling with an unknown block dependence structure under correlation constraints.
Journal of Computational Statistics and Data Analysis, 119:139–154.

References II

- Silverman, B. W. (1986).
  Density Estimation for Statistics and Data Analysis, volume 37.
- Soize, C. and Ghanem, R. (2017).
  Probabilistic learning on manifold for optimization under uncertainties.
  Proceeding of Uncecomp 2017, pages 1–15.
- van der Laan, M., Dudoit, S., and Keles, S. (2004).
  Asymptotic optimality of likelihood-based cross-validation.
  Statistical Applications in Genetics and Molecular Biology, 3(1).