(Non)linear dimension reduction of input parameter space using gradient information

Clémentine PRIEUR
Grenoble Alpes University

Journée Réduction de dimension
Groupe Statistique Mathématique de la SFdS
Institut Henri Poincaré, January 12, 2023

 Olivier Zahm (Université Grenoble Alpes, France)

 Paul Constantine (University of Colorado Boulder, USA)

The collaboration was initiated during the Opening Workshop of the QMC and High-Dimensional Sampling Methods for Applied Mathematics program (2017) of the Statistical and Applied Mathematical Sciences Institute (SAMSI).


Olivier Zahm (Université Grenoble Alpes, France)


Daniele Bigoni (formerly at MIT, USA)


Youssef Marzouk (MIT, USA)

This work was supported by the Inria associate team UNQUESTIONABLE (UNcertainty QUantification is ESenTIal for OceaNic & Atmospheric flow proBLEms).

⋆ Our framework is the following:

$$\mathcal{M} : \left\{ \begin{array}{rcl} \mathcal{X} = \prod_{i=1}^{d} \mathcal{X}_i & \rightarrow & \mathcal{Y} \\ \mathbf{x} & \mapsto & y = \mathcal{M}(x_1, \ldots, x_d) \end{array} \right. \quad \text{with}$$

- $\mathcal{M}$ expensive to evaluate,
- high dimension $d \gg 1$.

- Our framework is the following:

$$\mathcal{M} : \begin{cases} \mathcal{X} = \prod_{i=1}^{d} \mathcal{X}_i & \rightarrow & \mathcal{Y} \\ \mathbf{x} & \mapsto & y = \mathcal{M}(x_1, \ldots, x_d) \end{cases} \quad \text{with}$$

  - $\mathcal{M}$ expensive to evaluate,
  - high dimension $d \gg 1$.

- We aim to:
  - define a r (new) inputs, $r \leq d$ to build a surrogate for $\mathcal{M}$,
  - exploit gradient information when available (e.g., automatic differentiation, adjoint method).

- Our framework is the following:

$$\mathcal{M} : \begin{cases} \mathcal{X} = \prod_{i=1}^{d} \mathcal{X}_i & \rightarrow & \mathcal{Y} \\ \mathbf{x} & \mapsto & y = \mathcal{M}(x_1, \ldots, x_d) \end{cases} \quad \text{with}$$

  - $\mathcal{M}$ expensive to evaluate,
  - high dimension $d \gg 1$.

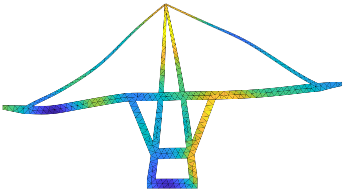
- We aim to:
  - define a r (new) inputs, $r \leq d$ to build a surrogate for $\mathcal{M}$,
  - exploit gradient information when available (e.g., automatic differentiation, adjoint method).

- More precisely, we seek for a decomposition of the form:

$$\mathcal{M}(x_1, \ldots, x_d) \approx f \circ g(\mathbf{x}) = f\left(g_1(x_1, \ldots, x_d), \ldots, g_r(x_1, \ldots, x_d)\right)$$

with $r \leq d$.

Illustration: resonance frequency of a bridge

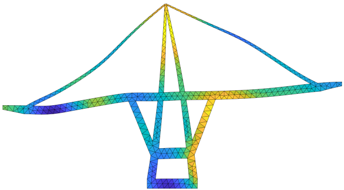


Parametrized eigenvalue problem

$$\mathcal{M}(\mathbf{x}) = \min_{v \in \mathbb{R}^{\mathcal{N}}} \frac{v^T K(\mathbf{x}) v}{v^T M v}$$

- $K(\mathbf{x})$: stiffness matrix, M: mass matrix
- $v \in \mathbb{R}^{\mathcal{N}}$, $\mathcal{N} = 960$ nodes in the finite element mesh
- the Young modulus field, $E(\mathbf{x}) = \exp\left(\sum_{i=1}^{d} x_i \sqrt{\sigma_i} \psi_i\right)$, with $\psi_i : \Omega \to \mathbb{R}$ and σ_i the i-th leading eigenfunctions and eigenvalues of kernel $c(s,t) = \sqrt{5}\exp(-\|s-t\|_2^2/20)$, is parametrized by $\mathbf{x} \sim \mathcal{N}_d(0, Id)$, $d = 32$.

Illustration: resonance frequency of a bridge



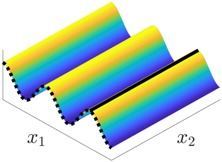
Parametrized eigenvalue problem

$$\mathcal{M}(\mathbf{x}) = \min_{v \in \mathbb{R}^{\mathcal{N}}} \frac{v^T K(\mathbf{x}) v}{v^T M v}$$

- $K(\mathbf{x})$: stiffness matrix, M: mass matrix
- $v \in \mathbb{R}^{\mathcal{N}}$, $\mathcal{N} = 960$ nodes in the finite element mesh
- the Young modulus field, $E(\mathbf{x}) = \exp\left(\sum_{i=1}^{d} x_i \sqrt{\sigma_i} \psi_i\right)$, with $\psi_i : \Omega \to \mathbb{R}$ and σ_i the i-th leading eigenfunctions and eigenvalues of kernel $c(s,t) = \sqrt{5} \exp(-\|s-t\|_2^2/20)$, is parametrized by $\mathbf{x} \sim \mathcal{N}_d(0, Id)$, $d = 32$.

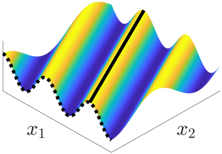
For this example, it is easy to compute model gradient $\nabla \mathcal{M}(\mathbf{x}) = (\partial_{x_1} \mathcal{M}(\mathbf{x}), \cdots, \partial_{x_d} \mathcal{M}(\mathbf{x}))$:

$$\partial_{x_i} \mathcal{M}(\mathbf{x}) = \frac{v(\mathbf{x})^T (\partial_{x_i} K(\mathbf{x})) v(\mathbf{x})}{v(\mathbf{x})^T M v(\mathbf{x})}, \text{ with } v(\mathbf{x}) = \operatorname*{argmin}_{v \in \mathbb{R}^{\mathcal{N}}} \frac{v^T K(\mathbf{x}) v}{v^T M v}.$$



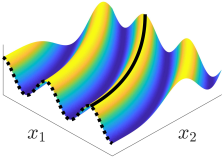
$\sin(x_1)$

$\underbrace{g(\mathbf{x}) = x_1}_{\text{linear in first canonical coordinate}}$



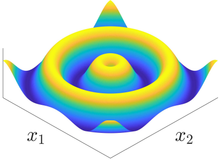
$\sin(x_1 + x_2)$

$\underbrace{g(\mathbf{x}) = x_1 + x_2}_{\text{linear}}$



$\sin(x_1 + x_2^2)$

$\underbrace{g(\mathbf{x}) = x_1 + x_2^2}_{\text{nonlinear}}$

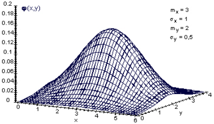


$\sin(x_1^2 + x_2^2)$

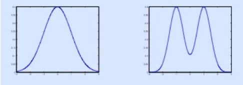
$\underbrace{g(\mathbf{x}) = x_1^2 + x_2^2}_{\text{nonlinear}}$

Uncertainty quantification framework

Uncertain input parameters are modeled by a probability distribution μ on $\mathcal{X}$, from experts' knowledge or from observations.

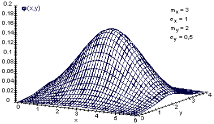


E.g., if the inputs are independent, this probability distribution is characterized by its marginals: $\mu(d\mathbf{x}) = \prod_{i=1}^{d} \mu_i(dx_i)$.

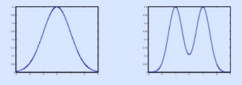


Uncertainty quantification framework

Uncertain input parameters are modeled by a probability distribution μ on $\mathcal{X}$, from experts' knowledge or from observations.



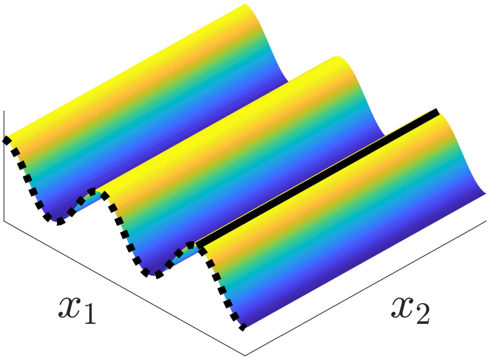
E.g., if the inputs are independent, this probability distribution is characterized by its marginals: $\mu(d\mathbf{x}) = \prod_{i=1}^{d} \mu_i(dx_i)$.



Approximation error is measured as

$$\mathbb{E}\left(\|\mathcal{M}(\mathbf{X}) - f \circ g(\mathbf{X})\|_{\mathcal{Y}}^{2}\right),$$

with some specific norm on $\mathcal{Y}$.


x_1
x_2

In the following,

$$\mathcal{M} : \begin{cases} \mathcal{X} = \mathbb{R}^d & \to & \mathcal{Y} = \mathbb{R}^p \\ \mathbf{x} & \mapsto & y = \mathcal{M}(x_1, \ldots, x_d) \end{cases}$$

For $p = 1$ (scalar output) and $\mathbf{u} \subset \{1, \ldots, d\}$, one defines the total Sobol' index for $\mathcal{M}$ associated to $\mathbf{u}$ as:

$$S_{\mathbf{u}}^{\text{tot}} = 1 - \frac{\mathrm{Var}\left[\mathbb{E}\left(Y|X_{-\mathbf{u}}\right)\right]}{\mathrm{Var}[Y]} = \frac{\mathbb{E}\left[\mathrm{Var}\left(Y|X_{-\mathbf{u}}\right)\right]}{\mathrm{Var}[Y]}$$

with $X_{-\mathbf{u}} = (X_i\ ,\ i \notin \mathbf{u})$ (see, e.g., Da Veiga et al. [2021]).

In the following,

$$\mathcal{M} : \begin{cases} \mathcal{X} = \mathbb{R}^d & \to & \mathcal{Y} = \mathbb{R}^p \\ \mathbf{x} & \mapsto & y = \mathcal{M}(x_1, \ldots, x_d) \end{cases}$$

For $p = 1$ (scalar output) and $\mathbf{u} \subset \{1, \ldots, d\}$, one defines the total Sobol' index for $\mathcal{M}$ associated to $\mathbf{u}$ as:

$$S_{\mathbf{u}}^{\text{tot}} = 1 - \frac{\mathrm{Var}\left[\mathbb{E}\left(Y|X_{-\mathbf{u}}\right)\right]}{\mathrm{Var}[Y]} = \frac{\mathbb{E}\left[\mathrm{Var}\left(Y|X_{-\mathbf{u}}\right)\right]}{\mathrm{Var}[Y]}$$

with $X_{-\mathbf{u}} = (X_i\ ,\ i \notin \mathbf{u})$ (see, e.g., Da Veiga et al. [2021]).

We then have the following equality Hart and Gremaud [2018]:

$$S_{\mathbf{u}}^{\text{tot}} = \frac{\|Y - \mathbb{E}\left(Y|X_{-\mathbf{u}}\right)\|^2}{\|Y - \mathbb{E}\left(Y\right)\|^2},$$

with $\|Y - \mathbb{E}\left(Y|X_{-\mathbf{u}}\right)\|^2 = \mathbb{E}\left(|\mathcal{M}(\mathbf{X}) - \mathbb{E}\left(Y|X_{-\mathbf{u}}\right)|^2\right)$.

From

$$S_{\mathbf{u}}^{\text{tot}} = \frac{\| Y - \mathbb{E}(Y|X_{-\mathbf{u}}) \|^2}{\| Y - \mathbb{E}(Y) \|^2},$$

with $\| Y - \mathbb{E}(Y|X_{-\mathbf{u}}) \|^2 = \mathbb{E}\left(|\mathcal{M}(\mathbf{X}) - \mathbb{E}(Y|X_{-\mathbf{u}})|^2\right)$, we deduce:

$$\begin{aligned} S_{\mathbf{u}}^{\text{tot}} \approx 0 \quad &\Leftrightarrow \quad \mathcal{M}(\mathbf{X}) \approx f(\mathbf{X}_{-\mathbf{u}}) \\ &\Leftrightarrow \quad \mathbf{X}_{\mathbf{u}} \text{ is useless to "explain" } Y = \mathcal{M}(\mathbf{X}) \end{aligned}$$

From

$$S_{\mathbf{u}}^{\text{tot}} = \frac{\| Y - \mathbb{E}(Y|X_{-\mathbf{u}}) \|^2}{\| Y - \mathbb{E}(Y) \|^2},$$

with $\| Y - \mathbb{E}(Y|X_{-\mathbf{u}}) \|^2 = \mathbb{E}\left(|\mathcal{M}(\mathbf{X}) - \mathbb{E}(Y|X_{-\mathbf{u}})|^2\right)$, we deduce:

$$\begin{aligned} S_{\mathbf{u}}^{\text{tot}} \approx 0 \quad &\Leftrightarrow \quad \mathcal{M}(\mathbf{X}) \approx f(\mathbf{X}_{-\mathbf{u}}) \\ &\Leftrightarrow \quad \mathbf{X}_{\mathbf{u}} \text{ is useless to "explain" } Y = \mathcal{M}(\mathbf{X}) \end{aligned}$$

Note that if $\mu(d\mathbf{x}) = \prod_{i=1}^{d} \mu_i(dx_i)$ then

$$S_{\mathbf{u}}^{\text{tot}} = \sum_{\mathbf{v} \subseteq \{1,\dots,d\}\,,\ \mathbf{u} \cap \mathbf{v} \neq \emptyset} S_{\mathbf{v}} \text{ and}$$

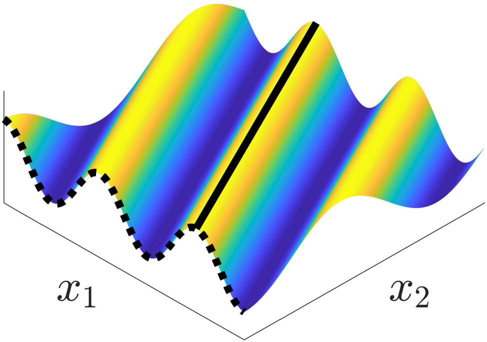
$$\begin{aligned} S_{\mathbf{u}}^{\text{tot}} \approx 0 \quad &\Leftrightarrow \quad \mathcal{M}(\mathbf{X}) \approx \mathcal{M}(\mathbf{x}_{\mathbf{u}}, \mathbf{X}_{-\mathbf{u}}) \text{ for } \textstyle\prod_{i \in \mathbf{u}} \mu_i\text{-almost all } \mathbf{x}_{\mathbf{u}} \\ &\Leftrightarrow \quad \mathbf{X}_{\mathbf{u}} \text{ is useless to "explain" } Y = \mathcal{M}(\mathbf{X}) \\ &\Leftrightarrow \quad \mathbf{X}_{\mathbf{u}} \text{ "can be fixed" to any value in the model} \end{aligned}$$

A natural extension to the vector-valued case:

$$S_{\mathbf{u}}^{\text{tot}} = \frac{\mathbb{E}(\|\mathcal{M}(\mathbf{X}) - \mathbb{E}(\mathcal{M}(\mathbf{X})|\mathbf{X}_{-\mathbf{u}})\|_{\mathcal{Y}}^2)}{\mathbb{E}(\|\mathcal{M}(\mathbf{X}) - \mathbb{E}(\mathcal{M}(\mathbf{X}))\|_{\mathcal{Y}}^2)},$$

with $\mathcal{Y} = \mathbb{R}^p$ endowed with a hilbertian norm $\|\cdot\|_{\mathcal{Y}}$ (see Lamboni et al. [2011], Gamboa et al. [2013], Zahm et al. [2020]).

Gradient based linear dimension reduction Constantine and Diaz [2017], Zahm et al. [2020]



Framework:

$$\mathbf{x} \in \mathbb{R}^d \mapsto \mathcal{M}(x_1, \ldots, x_d) \ \in \mathcal{Y}$$

with $\mathcal{Y} = \mathbb{R}^p$ endowed with a Hilbertian norm $\|\cdot\|_\mathcal{Y}$.

One aims at approximating $\mathcal{M}$ by a ridge function (a function which is constant along a subspace). More specifically, one seeks for $r \leq d$ and $A \in \mathbb{R}^{r\times d}$ such that:

$$\mathcal{M}(\mathbf{x}) \approx f(A\,\mathbf{x}) \text{ with } f : \mathbb{R}^r \to \mathcal{Y}\,,$$

or equivalently for $r \leq d$ and a rank-r projector $P_r \in \mathbb{R}^{d\times d}$ such that:

$$\mathcal{M}(\mathbf{x}) \approx h(P_r\,\mathbf{x}) \text{ with } h : \mathbb{R}^d \to \mathcal{Y}\,.$$

We assume $\mathbf{X} \sim \mu = \mathcal{N}(m, \Sigma)$.

Controlled approximation problem Given $\varepsilon \geq 0$, find r, h and a rank-r projector P_r such that

$$\mathbb{E}(\|\mathcal{M}(\mathbf{X}) - h(P_r \mathbf{X})\|_{\mathcal{Y}}^2) \leq \varepsilon.$$

Procedure:

1. derive an upper bound for the error
$$\|\mathcal{M} - h \circ P_r\| \leq \mathcal{R}(h, P_r)$$
2. fix r and solve
$$\min_{h, P_r} \mathcal{R}(h, P_r)$$
3. increase r until
$$\min_{h, P_r} \mathcal{R}(h, P_r) \leq \varepsilon$$

Note that P_r is not restricted to be a projector onto the canonical coordinates.

Derivation of the upper bound

For any projector P_r,

$$\|\mathcal{M} - \mathbb{E}_\mu(\mathcal{M}|\sigma(P_r))\| = \min_h \|\mathcal{M} - h \circ P_r\|.$$

From Poincaré type inequalities, we can deduce that for $\mathcal{M} : \mathbb{R}^d \to \mathcal{Y}$ smooth vector-valued and for any projector P_r,

$$\|\mathcal{M} - \mathbb{E}_\mu(\mathcal{M}|\sigma(P_r))\| \leq \sqrt{\mathrm{trace}\big(H(I_d - P_r)\Sigma(I_d - P_r)^T\big)}$$

with matrix $H \in \mathbb{R}^{d\times d}$ defined by

$$\boxed{H = \int (\nabla\mathcal{M})^*(\nabla\mathcal{M})\mathrm{d}\mu}$$

where

$$\begin{cases} \nabla\mathcal{M}(x) : \mathbb{R}^d \to \mathbb{R}^p \text{ Jacobian of } \mathcal{M} \text{ at } x \\ \nabla\mathcal{M}(x)^* \text{ is the adjoint of } \nabla\mathcal{M}(x) \end{cases}$$

What is the matrix H ?

$$H = \int (\nabla \mathcal{M})^* (\nabla \mathcal{M}) \mathrm{d}\mu \quad \in \mathbb{R}^{d \times d}$$

- Vector-valued case: $\mathcal{Y} = \mathbb{R}^p$ with $\|\cdot\|_{\mathcal{Y}}$ such that $\|v\|_{\mathcal{Y}}^2 = v^T R_{\mathcal{Y}} v$ for some SPD matrix $R_{\mathcal{Y}} \in \mathbb{R}^{p \times p}$. Then

$$H = \int (\nabla \mathcal{M})^T R_{\mathcal{Y}} (\nabla \mathcal{M}) \, \mathrm{d}\mu$$

with

$$\nabla \mathcal{M} = \begin{pmatrix} \frac{\partial \mathcal{M}_1}{\partial x_1} & \cdots & \frac{\partial \mathcal{M}_1}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathcal{M}_p}{\partial x_1} & \cdots & \frac{\partial \mathcal{M}_p}{\partial x_d} \end{pmatrix}$$

- Scalar-valued case: $\mathcal{Y} = \mathbb{R}$ with $\|\cdot\|_{\mathcal{Y}} = |\cdot|$, then

$$H = \int (\nabla \mathcal{M})(\nabla \mathcal{M})^T \, \mathrm{d}\mu$$

with

$$\nabla \mathcal{M} = \begin{pmatrix} \frac{\partial \mathcal{M}}{\partial x_1} \\ \vdots \\ \frac{\partial \mathcal{M}}{\partial x_d} \end{pmatrix}$$

$\rightsquigarrow\rightsquigarrow$ Active-Subspace method Constantine and Diaz [2017]

Minimizing the upper bound

Let (v_i, λ_i) be the i-th generalized eigenpair of (H, Σ^{-1}):

$$H v_i = \lambda_i \Sigma^{-1} v_i.$$

One has $\lambda_1 \geq \cdot \geq \lambda_i \geq \cdot \geq \lambda_d$ and

$$\min_{P_r} \sqrt{\text{trace}(H(I_d - P_r)\Sigma(I_d - P_r)^T)} = \sqrt{\sum_{i=r+1}^{d} \lambda_i}$$

A solution is the Σ^{-1}-orthogonal proj. P_r onto $\text{span}\{v_1, \ldots, v_r\}$, $P_r = \left(\sum_{i=1}^r v_i v_i^T\right) \Sigma^{-1}$, and

- a fast decay in λ_i ensures $\sqrt{\sum_{i=r+1}^d \lambda_i} \leq \varepsilon$ for $r = r(\varepsilon) \ll d$,
- H provides a test that reveals the low-effective dimension.

Let's come back to the upper bound, namely,

$$\|\mathcal{M} - \mathbb{E}_\mu(\mathcal{M}|\sigma(P_r))\| \leq \sqrt{\text{trace}(H(I_d - P_r)\Sigma(I_d - P_r)^T)}.$$

Choosing $\mathcal{Y} = \mathbb{R}^p$ and P_r as the projector that extracts the coordinates of $\mathbf{X}$ indexed by $\mathbf{u}$, we get:

$$S_{\mathbf{u}}^{\text{tot}} = \frac{\|\mathcal{M} - \mathbb{E}_\mu(\mathcal{M}|\sigma(I_d - P_r))\|^2}{\|\mathcal{M} - \mathbb{E}_\mu(\mathcal{M})\|^2}$$

thus

$$\begin{aligned} S_{\mathbf{u}}^{\text{tot}} &\leq \frac{\text{trace}\left(\Sigma P_r^T H P_r\right)}{\|\mathcal{M} - \mathbb{E}_\mu(\mathcal{M})\|^2} \\ &= \frac{\sum_{i\in\mathbf{u}} \mathrm{Var}(X_i) H_{i,i}}{\|\mathcal{M} - \mathbb{E}_\mu(\mathcal{M})\|^2}. \end{aligned}$$

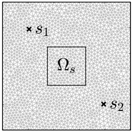
See, e.g., Sobol' & Kucherenko, 2009 and Lamboni *et al.*, 2013 for similar results in the case $p = 1$ (scalar output).

A numerical example

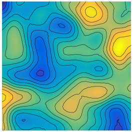
Diffusion problem on $\Omega = [0,1]^2$: $\begin{cases} \nabla \cdot \kappa \nabla u &= 0 & \text{in } \Omega \\ u &= x + y & \text{on } \partial\Omega \end{cases}$

- Random diffusion field κ, log-normal distribution.
- After finite element discretization:

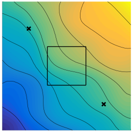
$$x = \log(\kappa) \in \mathbb{R}^{3252} \sim \mu = \mathcal{N}(0, \Sigma)$$



(a) mesh, 3252 elements



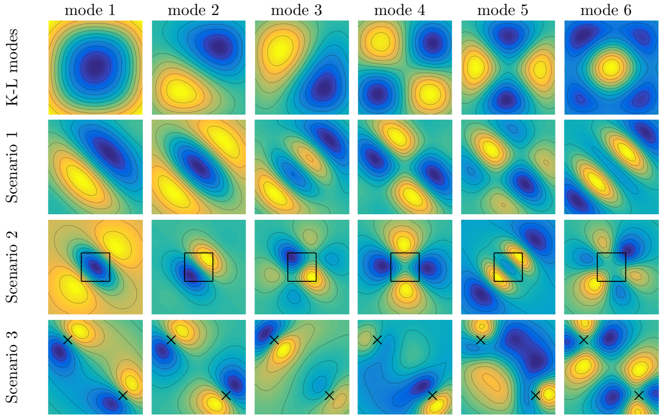
(b) log. diffusion field



(c) solution

1. Scenario 1 $\mathcal{M} : x \mapsto u \in \mathcal{Y} \subset H^1(\Omega)$, $p = 1691$ (number of nodes in the mesh for FEM);
2. Scenario 2 $\mathcal{M} : x \mapsto u_{|\Omega_s} \in \mathcal{Y} \subset H^1(\Omega_s)$, $p = 168$;
3. Scenario 3 $\mathcal{M} : x \mapsto (u_{|s_1}, u_{|s_2}) \in \mathcal{Y} = \mathbb{R}^2$ (canonical norm).

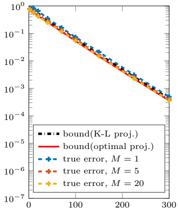
Modes $v_1, v_2, \ldots$



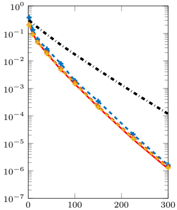
$$\text{Im}(P_r) = \text{span}\{v_1, v_2, \ldots, v_r\}$$

Approximation of the conditional expectation assuming H is known

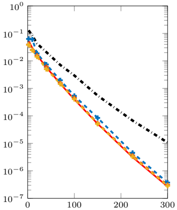
$$\mathbb{E}_\mu(\mathcal{M}|\sigma(P_r)) \;\approx\; \hat{F}_r : x \mapsto \frac{1}{M}\sum_{k=1}^{M} \mathcal{M}(P_r x + (I_d - P_r)\mathbf{Z}^{(k)}), \quad \mathbf{Z}^{(k)} \overset{iid}{\sim} \mu$$



$\mathcal{M} : x \mapsto u$



$\mathcal{M} : x \mapsto u_{|\Omega_s}$



$\mathcal{M} : x \mapsto (u_{|s_1}, u_{|s_2})$

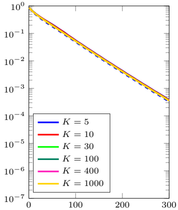
$$\|\mathcal{M} - \hat{F}_r\| = function(r)$$

We can show that

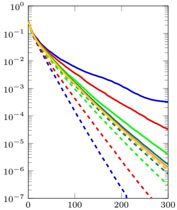
$$\mathbb{E}\left(\|\mathcal{M} - \hat{F}_r\|^2\right) \leq (1 + M^{-1})\ \mathrm{trace}(\Sigma(I_d - P_r^T)H(I_d - P_r))$$

Approximation of H to get the projector

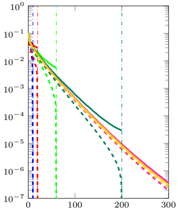
$$H \approx \widehat{H} = \frac{1}{K}\sum_{k=1}^{K} (\nabla \mathcal{M}(\mathbf{X}^{(k)}))^* (\nabla \mathcal{M}(\mathbf{X}^{(k)})), \quad \mathbf{X}^{(k)} \overset{iid}{\sim} \mu$$



$\mathcal{M} : x \mapsto u$



$\mathcal{M} : x \mapsto u_{|\Omega_s}$



$\mathcal{M} : x \mapsto \left(u_{|s_1}, u_{|s_2}\right)$

$\sqrt{\mathrm{trace}(\Sigma(I_d - \hat{P}_r^T)\widehat{H}(I_d - \hat{P}_r))} = \mathit{function}(r)$ (dashed curves)

$\sqrt{\mathrm{trace}(\Sigma(I_d - \hat{P}_r^T)H(I_d - \hat{P}_r))} = \mathit{function}(r)$ (solid curves)

Notice that $\mathrm{rank}(\widehat{H}) \leq K \ \max_{1\leq k\leq K} \mathrm{rank}\left(\nabla\mathcal{M}(\mathbf{X}^{(k)})\right) \leq K \dim(\mathcal{Y})$ (see also Zahm et al. [2022])

Beyond Gaussian uncertainty

Let $d\mu(x) \sim \exp\Big(- V(x) - \Psi(x) \Big) dx$. Assume

1. $\mathrm{supp}(\mu)$ **convex**,
2. (**Bakry–Émery** theorem) V a **convex** potential with
$$\nabla^2 V(x) \succeq \Gamma, \text{ with } \Gamma \text{ SPD matrix,}$$
3. (**Holley–Stroock** perturbation lemma) Ψ **bounded** with
$$\exp(\sup \Psi - \inf \Psi) \leq \kappa\,.$$

Then μ satisfies the subspace Poincaré inequality (Zahm et al. [2022]):

$$\|\mathcal{M} - \mathbf{E}[\mathcal{M}(\mathbf{X})|{P_r}^T \mathbf{X}]\|^2 \leq \kappa\, \mathrm{trace}[\Sigma(I_d - {P_r}^T) H (I_d - P_r))]$$

for any smooth function $\mathcal{M}$ and any projector P_r.

Beyond Gaussian uncertainty

Let $d\mu(x) \sim \exp\Big(-V(x) - \Psi(x)\Big)dx$. Assume

1. $\mathrm{supp}(\mu)$ **convex**,
2. (**Bakry–Émery** theorem) V a **convex** potential with
$$\nabla^2 V(x) \succeq \Gamma, \text{ with } \Gamma \text{ SPD matrix},$$
3. (**Holley–Stroock** perturbation lemma) Ψ **bounded** with
$$\exp(\sup \Psi - \inf \Psi) \leq \kappa\,.$$

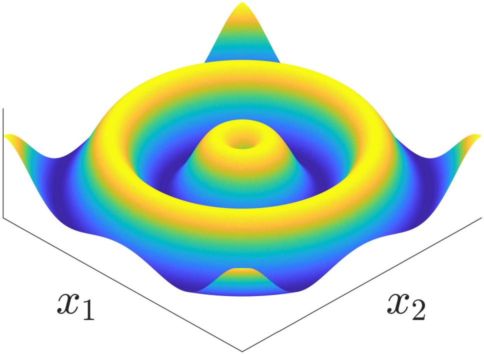
Then μ satisfies the subspace Poincaré inequality (Zahm et al. [2022]):

$$\|\mathcal{M} - \mathbf{E}[\mathcal{M}(\mathbf{X})|{P_r}^T\mathbf{X}]\|^2 \leq \kappa\,\mathrm{trace}[\Sigma(I_d - {P_r}^T)H(I_d - P_r))]$$

for any smooth function $\mathcal{M}$ and any projector P_r.

- Gaussian mixtures,
- uniform measures on compact & convex sets
- any measure such that $d\mu(\mathbf{x}) \geq \alpha > 0$ on compact & convex sets.

Extension to nonlinear dimension reduction Bigoni et al. [2022]



$$\mathcal{M} : \begin{cases} \mathcal{X} \subset \mathbb{R}^d & \rightarrow & \mathbb{R} \\ \mathbf{x} & \mapsto & y = \mathcal{M}(x_1, \ldots, x_d) \end{cases}$$

$$\mathcal{M}(x_1, \ldots, x_d) \approx f \circ g(\mathbf{x}) = f\left(g_1(x_1, \ldots, x_d), \ldots, g_r(x_1, \ldots, x_d)\right),$$

with the feature map g is not necessarily linear.

We propose, for any $r \leq d$, a two-step procedure.

- Step 1, construction of the feature map g:
  solve $\min_{g \in \mathcal{G}_r} J(g_1, \ldots, g_r)$ with J a gradient-based cost function.

- Step 2, construction of the profile fucntion f:

$$\text{solve} \quad \min_{f \in \mathcal{F}_r} \mathbb{E}\left[\left(\mathcal{M}(\mathbf{X}) - f \circ g(\mathbf{X})\right)^2\right].$$

Choice of the cost function J

Note that, if $\mathcal{M}(x_1, \ldots, x_d) = f \circ g(\mathbf{x})$, then

$$\nabla\mathcal{M}(\mathbf{x}) = \underbrace{\nabla g(\mathbf{x})^T}_{\in\mathbb{R}^{d\times r}} \underbrace{\nabla f(g(\mathbf{x}))}_{\in\mathbb{R}^{r}} \Rightarrow \nabla\mathcal{M}(\mathbf{x}) \in \mathrm{range}(\nabla g(x)^T).$$

A natural choice for J is then

$$J(g) := \mathbb{E}\left[\left\|\nabla\mathcal{M}(\mathbf{X}) - \Pi_{\mathrm{range}(\nabla g(\mathbf{X})^T)}\nabla\mathcal{M}(\mathbf{X})\right\|^2\right].$$

Choice of the cost function J

Note that, if $\mathcal{M}(x_1,\dots,x_d) = f\circ g(\mathbf{x})$, then

$$\nabla\mathcal{M}(\mathbf{x}) = \underbrace{\nabla g(\mathbf{x})^T}_{\in\mathbb{R}^{d\times r}}\underbrace{\nabla f(g(\mathbf{x}))}_{\in\mathbb{R}^r} \Rightarrow \nabla\mathcal{M}(\mathbf{x}) \in \mathrm{range}(\nabla g(x)^T).$$

A natural choice for J is then

$$J(g) := \mathbb{E}\left[\left\|\nabla\mathcal{M}(\mathbf{X}) - \Pi_{\mathrm{range}(\nabla g(\mathbf{X})^T)}\nabla\mathcal{M}(\mathbf{X})\right\|^2\right].$$

We have proven $\mathcal{M} = f\circ g \Rightarrow J(g) = 0$. Question a) Is the reciprocal true?

Question a): is the reciprocal $\Uparrow$ true? yes!

Proposition:

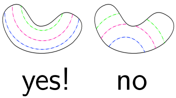
Assume $\mathcal{M} \in \mathcal{C}^1(\mathcal{X};\mathbb{R})$ and $g \in \mathcal{G}_r \subset \mathcal{C}^1(\mathcal{X};\mathbb{R}^r)$.
Assume that the level-sets of g are such that

$$g^{-1}(\{\mathbf{z}\}) = \{\mathbf{x} \in \mathcal{X} : g(\mathbf{x}) = \mathbf{z}\},$$

are **pathwise-connected** for any $z \in \mathbb{R}^r$. Then

$$J(g) = 0 \Rightarrow \exists\, f \text{ such that } \mathcal{M} = f \circ g$$

Are g's level sets pathwise-connected?



Examples of feature maps $g : \mathcal{X} \to \mathbb{R}$ with $\mathcal{X}$ convex and with smoothly pathwise connected level-sets:

Affine feature map Any function $g(\mathbf{x}) = A\mathbf{x} + b$ with $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$;

Feature map following from a C^1-diffeomorphism Any function $g(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_m(\mathbf{x}))$ where $\phi_i(\mathbf{x})$ is the i-th component of $\phi(\mathbf{x})$, with $\phi : \mathcal{X} \to \mathcal{X}$ a C^1-diffeomorphism;

Polynomial feature map Any polynomial function on $\mathcal{X} = \mathbb{R}^d$ such that for all $\mathbf{z} \in g(\mathcal{X})$, the zeros of the polynomial $\mathbf{x} \mapsto g(\mathbf{x}) - \mathbf{z}$ are pathwise-connected.

Examples of feature maps $g : \mathcal{X} \to \mathbb{R}$ with $\mathcal{X}$ convex and with smoothly pathwise connected level-sets:

Affine feature map Any function $g(\mathbf{x}) = A\mathbf{x} + b$ with $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$;

Feature map following from a C^1-diffeomorphism Any function $g(\mathbf{x}) = (\phi_1(\mathbf{x}), \ldots, \phi_m(\mathbf{x}))$ where $\phi_i(\mathbf{x})$ is the i-th component of $\phi(\mathbf{x})$, with $\phi : \mathcal{X} \to \mathcal{X}$ a C^1-diffeomorphism;

Polynomial feature map Any polynomial function on $\mathcal{X} = \mathbb{R}^d$ such that for all $\mathbf{z} \in g(\mathcal{X})$, the zeros of the polynomial $\mathbf{x} \mapsto g(\mathbf{x}) - \mathbf{z}$ are pathwise-connected.Computing the number of connected components (i.e., the zeroth Betti number) of an algebraic set like $\{\mathbf{x} : g(\mathbf{x}) - \mathbf{z}\}$ is a difficult question, commonly encountered in algebraic geometry.

Question b): does $J(g) \approx 0$ implies $\mathcal{M} \approx f \circ g$? yes!

Denote by $\mathbb{C}(Z)$ the **Poincaré constant** of a random vector Z, that is, the smallest constant such that

$$\mathrm{Var}(h(Z)) \leq \mathbb{C}(Z)\,\mathbb{E}\left[\left\|\nabla h(Z)\right\|^2\right]$$

holds for any smooth function $h : \mathrm{supp}(Z) \to \mathbb{R}$.

Proposition:

Assume $\mathcal{G}_r \subset \mathcal{C}^1(\mathbf{X}; \mathbb{R}^r)$ and $\mathrm{rank}\left(\nabla g(\mathbf{x})^T\right) = r\ \forall\, g \in \mathcal{G}_r$, $\forall\, \mathbf{x} \in \mathcal{X}$. Assume

$$\mathbb{C}(\mathbf{X}|\mathcal{G}_r) := \sup_{g\in\mathcal{G}_r} \sup_{\mathbf{z}\in g(\mathcal{X})} \mathbb{C}(\mathbf{X}|g(\mathbf{X}) = \mathbf{z}) < \infty.$$

Then for any $g \in \mathcal{G}_r$, there exists a profile $f : \mathbb{R}^r \to \mathbb{R}$ such that

$$\mathbb{E}\left[\left(\mathcal{M}(\mathbf{X}) - f \circ g(\mathbf{X})\right)^2\right] \leq \mathbb{C}(\mathbf{X}|\mathcal{G}_r)\, J(g).$$

Example: if $\mathcal{G}_r = \{\mathbf{x} \mapsto U^T\mathbf{x} : U \in \mathbb{R}^{d\times r} \text{ orth. columns}\}$ and if $\mathbf{X} \sim \mathcal{N}(0, I_d)$, then

$$\mathbb{C}(\mathbf{X}|\mathcal{G}_r) = 1$$

Although assuming $\mathbb{C}(\mathbf{X}|\mathcal{G}_r) < \infty$ is usual, e.g., in the analysis of Markov semigroups or in molecular dynamics, proving it remains an open challenge in more general settings.

Question c): how to minimize $g \mapsto J(g)$? We seek for g solving

$$\min_{g=(g_1,\dots,g_r)\in\mathcal{G}_r} J(g) = \mathbb{E}\left[\left\|\nabla\mathcal{M}(\mathbf{X}) - \Pi_{\mathrm{range}(\nabla g(\mathbf{X})^T)}\nabla\mathcal{M}(\mathbf{X})\right\|^2\right]$$

with $\mathcal{G}_r = \mathcal{G}^r = \mathrm{span}\{\Phi_1,\dots,\Phi_K\}^r$.

Question c): how to minimize $g \mapsto J(g)$? We seek for g solving

$$\min_{g=(g_1,\dots,g_r)\in\mathcal{G}_r} J(g) = \mathbb{E}\left[\left\|\nabla\mathcal{M}(\mathbf{X}) - \Pi_{\mathrm{range}(\nabla g(\mathbf{x})^T)}\nabla\mathcal{M}(\mathbf{X})\right\|^2\right]$$

with $\mathcal{G}_r = \mathcal{G}^r = \mathrm{span}\{\Phi_1,\dots,\Phi_K\}^r$.

It is equivalent to seek for g solving

$$\max_{G\in\mathbb{R}^{\#\mathcal{G}\times r}} \mathcal{R}(G) = \mathbb{E}\left[\mathrm{trace}(G^T H(\mathbf{X})G)(G^T\Sigma(\mathbf{X})G)^{-1}\right] \text{ where}$$

$$H(\mathbf{x}) = \nabla\Phi(\mathbf{x})(\nabla\mathcal{M}(\mathbf{x})\nabla\mathcal{M}(\mathbf{x})^T)\nabla\Phi(\mathbf{x})^T,$$
$$\Sigma(\mathbf{x}) = \nabla\Phi(\mathbf{x})\nabla\Phi(\mathbf{x})^T, \text{ with } \Phi(\mathbf{x}) = (\Phi_1(\mathbf{x}),\dots,\Phi_K(\mathbf{x})).$$

Maximization is solved with a quasi-Newton algorithm.

For linear feature maps, $g(\mathbf{x}) = A\,\mathbf{x}$, our procedure coincides with active subspace method.

Adaptive construction of g from $\{\mathbf{X}^{(i)}, \mathcal{M}(\mathbf{X}^{(i)}), \nabla\mathcal{M}(\mathbf{X}^{(i)})\}_{i=1}^{N}$

Empirical cost

We first replace $\mathcal{R}(G)$ by its empirical counterpart:

$$\hat{\mathcal{R}}^N(G) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{trace}(G^T H(\mathbf{X}^{(i)}) G)(G^T \Sigma(\mathbf{X}^{(i)}) G)^{-1}.$$

For any $1 \leq r \leq d$, we adapt the complexity of $\mathcal{G}_r = \mathcal{G}^r$ to the sample size N.

Matching Pursuit

We use a state-of-the-art Migliorati [2015, 2019] reduced-set matching pursuit algorithm on downward-closed polynomial spaces to build g.

Cross Validation

is used to know when to stop the iterations (before it overfits).

More precisely, to adapt the complexity of $\mathcal{G}$ with respect to the sample size N, one uses the following tools:

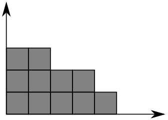
More precisely, to adapt the complexity of $\mathcal{G}$ with respect to the sample size N, one uses the following tools:

Downward closed polynomial spaces

$$\mathcal{G} = \mathbb{P}_\Lambda[\mathbb{R}^d] = \text{span}\{x_1^{\nu_1} \dots x_d^{\nu_d}, \nu \in \Lambda\}$$

where $\Lambda \subset \mathbb{N}^d$ is a downward closed set, that is:

$$\nu \in \Lambda \text{ and } \mu \leq \nu \quad \Rightarrow \quad \mu \in \Lambda$$



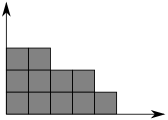
More precisely, to adapt the complexity of $\mathcal{G}$ with respect to the sample size N, one uses the following tools:

Downward closed polynomial spaces

$$\mathcal{G} = \mathbb{P}_\Lambda[\mathbb{R}^d] = \text{span}\{x_1^{\nu_1}\dots x_d^{\nu_d}, \nu \in \Lambda\}$$

where $\Lambda \subset \mathbb{N}^d$ is a downward closed set, that is:

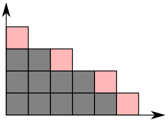
$$\nu \in \Lambda \text{ and } \mu \leq \nu \quad \Rightarrow \quad \mu \in \Lambda$$



Matching Pursuit

$$\Lambda_{k+1} = \Lambda_k \cup \{\nu_{k+1}\}$$

$$\nu_{k+1} \in \underset{\nu \in \text{ReducedMargin}(\Lambda_k)}{\text{argmax}} |\partial_\nu \hat{\mathcal{R}}^N(G_k^*)|$$



where G_k^* is the minimizer of $\hat{\mathcal{R}}^N(\cdot)$ over Λ_k.

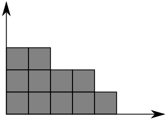
More precisely, to adapt the complexity of $\mathcal{G}$ with respect to the sample size N, one uses the following tools:

Downward closed polynomial spaces

$$\mathcal{G} = \mathbb{P}_\Lambda[\mathbb{R}^d] = \text{span}\{x_1^{\nu_1} \dots x_d^{\nu_d}, \nu \in \Lambda\}$$

where $\Lambda \subset \mathbb{N}^d$ is a downward closed set, that is:

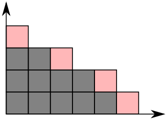
$$\nu \in \Lambda \text{ and } \mu \leq \nu \quad \Rightarrow \quad \mu \in \Lambda$$



Matching Pursuit

$$\Lambda_{k+1} = \Lambda_k \cup \{\nu_{k+1}\}$$

$$\nu_{k+1} \in \underset{\nu \in \text{ReducedMargin}(\Lambda_k)}{\text{argmax}} |\partial_\nu \hat{\mathcal{R}}^N(G_k^*)|$$



where G_k^* is the minimizer of $\hat{\mathcal{R}}^N(\cdot)$ over Λ_k.

Cross Validation

To know when to stop the iterations (before it overfits).

Once g is computed, how to construct f?

$$\min_{f\in\mathcal{F}_r} \frac{1}{N}\sum_{i=1}^{N}\left(\mathcal{M}(\mathbf{X}^{(i)}) - f\circ g(\mathbf{X}^{(i)})\right)^2 \underbrace{+\|\nabla\mathcal{M}(\mathbf{X}^{(i)}) - \nabla f\circ g(\mathbf{X}^{(i)})\|^2}_{\text{recycle the gradients}}$$

As for $\mathcal{G}$, we adapt the complexity of $\mathcal{F}_r = \mathcal{F}^r$ using reduced-set matching pursuit algorithm on downward-closed polynomial spaces.

Once g is computed, how to construct f?

$$\min_{f\in\mathcal{F}_r} \frac{1}{N}\sum_{i=1}^{N} \left(\mathcal{M}(\mathbf{X}^{(i)}) - f\circ g(\mathbf{X}^{(i)})\right)^2 \underbrace{+\|\nabla\mathcal{M}(\mathbf{X}^{(i)}) - \nabla f\circ g(\mathbf{X}^{(i)})\|^2}_{\text{recycle the gradients}}$$

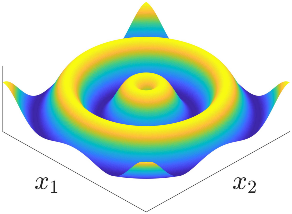
As for $\mathcal{G}$, we adapt the complexity of $\mathcal{F}_r = \mathcal{F}^r$ using reduced-set matching pursuit algorithm on downward-closed polynomial spaces.

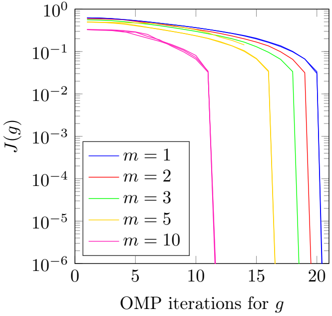
Benchmark algorithm (without dimension reduction):

$$\min_{v\in\mathcal{V}} \frac{1}{N}\sum_{i=1}^{N} \left(\mathcal{M}(\mathbf{X}^{(i)}) - v(\mathbf{X}^{(i)})\right)^2 \underbrace{+\|\nabla\mathcal{M}(\mathbf{X}^{(i)}) - \nabla v(\mathbf{X}^{(i)})\|^2}_{\text{recycle the gradients}}$$

Illustration: isotropic function

$$\mathcal{M}(\mathbf{x}) = \cos\left(\sqrt{x_1^2 + \ldots + x_d^2}\right)$$
$$\mu = \mathcal{N}(0, I_d)$$
$$\mathbf{x} \in \mathbb{R}^{20}$$
$$N = 100$$





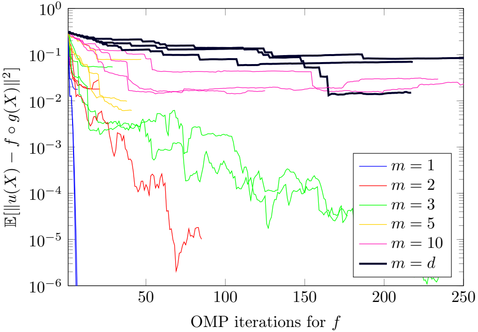
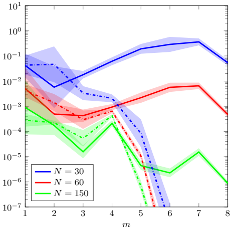
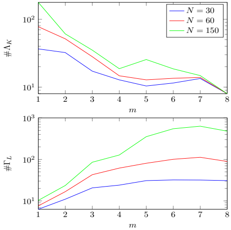


Illustration: Borehole function

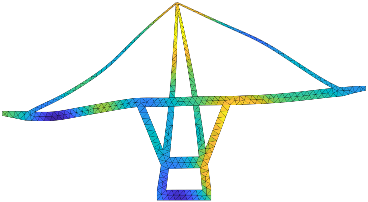
$$\mathcal{M}(\mathbf{x}) = \frac{2\pi T_{\mathcal{M}}(H_{\mathcal{M}} - H_\ell)}{\ln(r/r_\omega)\left(1 + \frac{2LT_{\mathcal{M}}}{\ln(r/r_\omega) r_\omega{}^2 K_\omega} + \frac{T_{\mathcal{M}}}{T_\ell}\right)}, \quad \begin{cases} x_1 = r_\omega & \sim \mathcal{N}(0.1, 3 \cdot 10^{-4}) \\ x_2 = r & \sim \log\mathcal{N}(7.71, 1.0112) \\ x_3 = T_{\mathcal{M}} & \sim \mathcal{U}(63\,070, 115\,600) \\ x_4 = H_{\mathcal{M}} & \sim \mathcal{U}(990, 1110) \\ x_5 = T_\ell & \sim \mathcal{U}(63.1, 116) \\ x_6 = H_\ell & \sim \mathcal{U}(700, 820) \\ x_7 = L & \sim \mathcal{U}(1120, 1\,680) \\ x_8 = K_\omega & \sim \mathcal{U}(9\,855, 12\,045) \end{cases}$$





Continuous lines: mean squared error $\mathbb{E}[(\mathcal{M}(\mathbf{X}) - f \circ g(\mathbf{X}))^2]$, Dashed lines: cost function $J(g)$. The width of the shaded region corresponds to the standard deviation over 20 experiments.

Illustration: resonance frequency of a bridge



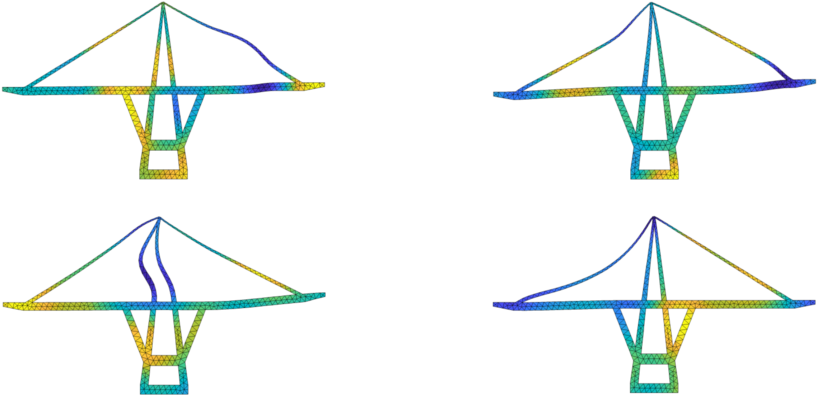
Parametrized eigenvalue problem

$$\mathcal{M}(\mathbf{x}) = \min_{v\in\mathbb{R}^{\mathcal{N}}} \frac{v^T K(\mathbf{x}) v}{v^T M v}$$

- $K(\mathbf{x})$: stiffness matrix
- M: mass matrix
- $v \in \mathbb{R}^{\mathcal{N}}$, $\mathcal{N} = 960$ nodes in the finite element mesh
- $\mathbf{x} \in \mathbb{R}^d$: Young modulus field ($d = 32$ KL modes)
- $N = 100$ (20 trials)

For this example, it is easy to compute model gradient $\nabla\mathcal{M}(\mathbf{x}) = (\partial_{x_1}\mathcal{M}(\mathbf{x}), \cdots, \partial_{x_d}\mathcal{M}(\mathbf{x}))$:

$$\partial_{x_i}\mathcal{M}(\mathbf{x}) = \frac{v(\mathbf{x})^T (\partial_{x_i} K(\mathbf{x})) v(\mathbf{x})}{v(\mathbf{x})^T M v(\mathbf{x})}, \text{ with } v(\mathbf{x}) = \operatorname*{argmin}_{v\in\mathbb{R}^{\mathcal{N}}} \frac{v^T K(\mathbf{x}) v}{v^T M v}.$$



Resonance frequency of a bridge. Four realizations of the Young modulus field **X** (color of the elements) and the associated resonance mode $v(\mathbf{X})$ (displacement of the mesh).

Results:

	$r=1$	$r=2$	$r=3$	$r=4$	$r=6$	$r=8$	$r=16$	$r=32$
Mean $\times 10^{12}$	1.6	1.5	1.1	1.2	1.3	1.5	1.6	1.4
Std $\times 10^{12}$	0.80	0.69	0.22	0.24	0.28	0.83	0.39	0.43
$\#\Lambda_K$	148 ($\pm$64)	129 ($\pm$45)	91 ($\pm$21)	80 ($\pm$23)	64 ($\pm$16)	57 ($\pm$9)	51 ($\pm$1)	32 ($\pm$0)
$\#\Gamma_L$	5 ($\pm$1)	8 ($\pm$1)	11 ($\pm$1)	15 ($\pm$3)	24 ($\pm$7)	44 ($\pm$24)	133 ($\pm$102)	102 ($\pm$70)

Mean and standard deviation of mean squared error $\mathbf{E}[(\mathcal{M}(\mathbf{X}) - f \circ g(\mathbf{X}))^2]$ over 20 experiments, where g and f are constructed adaptively with $N = 100$ samples. Mean squared error is computed on a (fixed) validation set of size 1000. The last two lines give mean($\pm$ std) of the cardinality of $\#\Lambda_K$ and $\#\Gamma_L$, which represent the complexity of g and f, respectively.

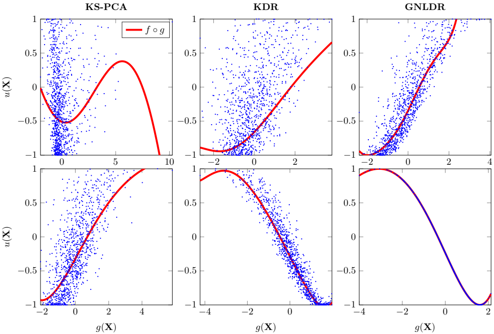
Comparison with nonlinear (NL) kernel supervised PCA and NL kernel dimension reduction.

$$\mathbf{Y} = \begin{pmatrix} \mathcal{M}(\mathbf{X}) \\ \nabla\mathcal{M}(\mathbf{X}) \end{pmatrix} \in \mathbb{R}^{1+d}.$$

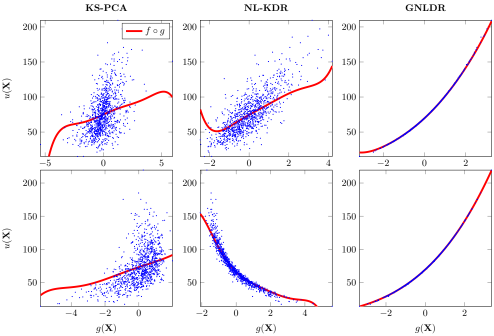
Kernel supervised PCA Barshan et al. [2011] aims to maximize the dependence between $G^T\Phi(\mathbf{X})$ and $\mathbf{Y}$ measured with the Hilbert-Schmidt norm of the cross-covariance operator restricted to an arbitrary reproducing kernel Hilbert space (RKHS).

Kernel dimension reduction Fukumizu et al. [2009] aims to minimize the dependence between $\mathbf{Y}$ and $\mathbf{Y}|G^T\Phi(\mathbf{X})$ measured with the Hilbert-Schmidt norm of the conditional covariance operator restricted to some RKHS.

In our experiments, we used squared exponential kernels for both $\kappa_{\mathbf{X}}$ and $\kappa_{\mathbf{Y}}$.



Isotropic function. Comparison of KS-PCA and NL-KDR with our method (GNLDR) for $m = 1$. Blue points: 1000 samples of $(g(\mathbf{X}), \mathcal{M}(\mathbf{X}))$. Red lines: function $g(\mathbf{x}) \mapsto f \circ g(\mathbf{x})$ with either $N = 50$ (top row) or $N = 500$ (bottom row). Here, f is a univariate polynomial of degree 6 and g a multivariate polynomial of degree 2.



Borehole function. Comparison of KS-PCA and NL-KDR with our method (GNLDR) for $m = 1$. Blue points: 1000 samples of $(g(\mathbf{X}), \mathcal{M}(\mathbf{X}))$. Red lines: function $g(\mathbf{x}) \mapsto f \circ g(\mathbf{x})$ with either $N = 30$ (top row) or $N = 300$ (bottom row). Here, f is a univariate polynomial of degree 6 and g a multivariate polynomial of degree 2.

Conclusion

- In this talk, we presented a trip around global sensitivity analysis (via total Sobol' indices) and (non)linear dimension reduction.
- We proposed a two-step algorithm to build the approximation $\mathcal{M}(\mathbf{x}) \approx f \circ g(\mathbf{x})$ adaptively with respect to the input/output sample. This algorithme takes into account gradient information.

Conclusion

- In this talk, we presented a trip around global sensitivity analysis (via total Sobol' indices) and (non)linear dimension reduction.
- We proposed a two-step algorithm to build the approximation $\mathcal{M}(\mathbf{x}) \approx f \circ g(\mathbf{x})$ adaptively with respect to the input/output sample. This algorithme takes into account gradient information.

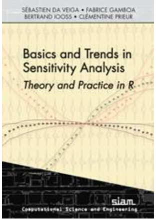
Perspectives

- It would be interesting to propose an optimal (or at least a clever) sampling procedure.
- Beyond polynomial approximation?
- Although assuming $\mathbb{C}(\mathbf{X}|\mathcal{G}_r) < \infty$ is usual, proving it remains an open challenge. Is it possible to choose the approximation class $\mathcal{G}_r$ such that $\mathbb{P}_{\mathbf{X}|\mathcal{G}_r}$ is the push-forward measure of the standard normal distribution through a Lipschitz map.
- ...

Thanks for your attention!

Thanks for your attention!

And a little bit of advertisement



Some references I

Elnaz Barshan, Ali Ghodsi, Zohreh Azimifar, and Mansoor Zolghadri Jahromi. Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. *Pattern Recognition*, 44(7):1357–1371, 2011.

Daniele Bigoni, Youssef Marzouk, Clémentine Prieur, and Olivier Zahm. Nonlinear dimension reduction for surrogate modeling using gradient information. *Information and Inference: A Journal of the IMA*, 05 2022. ISSN 2049-8772. doi: 10.1093/imaiai/iaac006. URL https://doi.org/10.1093/imaiai/iaac006.

Paul G Constantine and Paul Diaz. Global sensitivity metrics from active subspaces. *Reliability Engineering & System Safety*, 162:1–13, 2017.

Sébastien Da Veiga, Fabrice Gamboa, Bertrand Iooss, and Clémentine Prieur. *Basics and Trends in Sensitivity Analysis: Theory and Practice in R*. SIAM, 2021.

Kenji Fukumizu, Francis R Bach, and Michael I Jordan. Kernel dimension reduction in regression. *The Annals of Statistics*, 37(4):1871–1905, 2009.

Fabrice Gamboa, Alexandre Janon, Thierry Klein, and Agnès Lagnoux. Sensitivity indices for multivariate outputs. *Comptes Rendus. Mathématique*, 351(7-8): 307–310, April 2013. URL https://hal.inria.fr/hal-00800847.

Some references II

J.L. Hart and P.A. Gremaud. An approximation theoretic perspective of sobol' indices with dependent variables. *International Journal for Uncertainty Quantification*, 8 (6), 2018.

M. Lamboni, H. Monod, and D. Makowski. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering and System Safety*, 96(4):450–459, 2011.

Giovanni Migliorati. Adaptive polynomial approximation by means of random discrete least squares. In *Numerical Mathematics and Advanced Applications-ENUMATH 2013*, pages 547–554. Springer, 2015.

Giovanni Migliorati. Adaptive approximation by optimal weighted least-squares methods. *SIAM Journal on Numerical Analysis*, 57(5):2217–2245, 2019.

I. M. Sobol' and S. Kucherenko. Derivative based global sensitivity measures and the link with global sensitivity indices. *Mathematics and Computers in Simulation*, 79: 3009–3017, 2009.

Olivier Zahm, Paul G Constantine, Clémentine Prieur, and Youssef M Marzouk. Gradient-based dimension reduction of multivariate vector-valued functions. *SIAM Journal on Scientific Computing*, 42(1):A534–A558, 2020.

Some references III

Olivier Zahm, Tiangang Cui, Kody Law, Alessio Spantini, and Youssef Marzouk. Certified dimension reduction in nonlinear bayesian inverse problems. *Mathematics of Computation*, 91(336):1789–1835, 2022.