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

Clémentine PRIEUR
Grenoble Alpes University

SAMO, March 14-16, 2022
Tallahassee, Florida (USA)

 Olivier Zahm (Université Grenoble Alpes, France)

 Daniele Bigoni (formerly at MIT, USA)

 Youssef Marzouk (MIT, USA)

 Paul Constantine (University of Colorado Boulder, USA)

⋆ Our framework is the following:

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

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

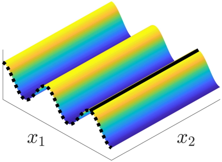
⋆ We aim to:

- select a subset of inputs to build a surrogate for $\mathcal{M}$,
- exploit gradient information when available (e.g., automatic differentiation, adjoint method).

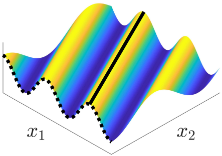
⋆ More precisely, we seek for a decomposition of the form:

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

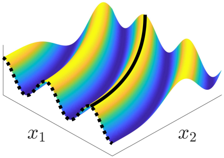
with $r \leq d$.



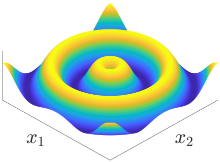
$\sin(x_1)$
$\underbrace{g(\mathbf{x}) = x_1}_{\text{linear in first canonical coordinate}}$



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



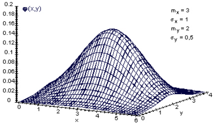
$\sin(x_1 + x_2^2)$
$\underbrace{g(\mathbf{x}) = x_1 + x_2^2}_{\text{nonlinear}}$



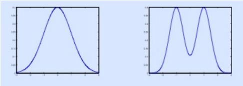
$\sin(x_1^2 + x_2^2)$
$\underbrace{g(\mathbf{x}) = x_1^2 + x_2^2}_{\text{nonlinear}}$

Uncertainty quantification framework

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



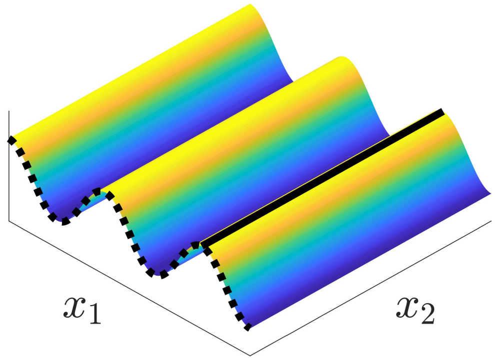
E.g., if the inputs are independent, this probability distribution is characterized by its marginals: $\mu(d\mathbf{x}) = \prod_{i=1}^{d} \mu_i(dx_i)$.



Approximation error is measured as

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

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



x_1
x_2

In the following,

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

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]:

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

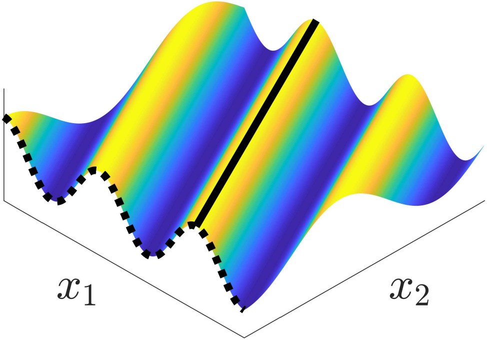
with $\|\, Y - \mathbb{E}\left(Y|X_{-\mathbf{u}}\right) \|^2 = \mathbb{E}\left(|\mathcal{M}(\mathbf{X}) - \mathbb{E}\left(Y|X_{-\mathbf{u}}\right)|^2\right)$.

A natural extension to the vector-valued [Zahm et al., 2020] case:

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

with V a vectorial Hilbert space and $\|\cdot\|_V$ the associated norm.

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



Framework:

$$\mathbf{x} \mapsto \mathcal{M}(x_1, \ldots, x_d) \; \in V$$

with $V = \mathbb{R}^p$ endowed with a Hilbertian norm $\| \cdot \|_V$.

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:

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

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

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

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

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

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

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 $\mathcal{M} : \mathbb{R}^d \to V$ smooth vector-valued and for any projector P_r,

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

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

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

where

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

What is the matrix H ?

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

- Algebraic case: $V = \mathbb{R}^p$ with $\|\cdot\|_V$ such that $\|v\|_V^2 = v^T R_V v$ for some SPD matrix $R_V \in \mathbb{R}^{p \times p}$. Then

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

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: $V = \mathbb{R}$ with $\|\cdot\|_V = |\cdot|$, 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}):

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

One has:

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

A solution is a Σ^{-1}-orthogonal proj. onto $\mathrm{span}\{v_1, \dots, v_r\}$ 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,

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

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

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

thus

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

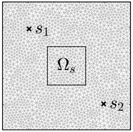
See, e.g., Sobol' & Kucherenko, 2009 and Lamboni *et al.*, 2013 for similar results in the case $p = 1$ (scalar output).

A numerical example

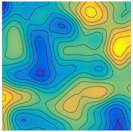
Diffusion problem on $\Omega = [0,1]^2$: $\begin{cases} \nabla \cdot \kappa \nabla u &= 0 & \text{in } \Omega \\ u &= x + y & \text{on } \partial\Omega \end{cases}$

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

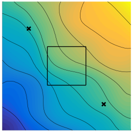
$$x = \log(\kappa) \in \mathbb{R}^{3252} \sim \mu = \mathcal{N}(0, \Sigma)$$



(a) mesh, 3252 elements



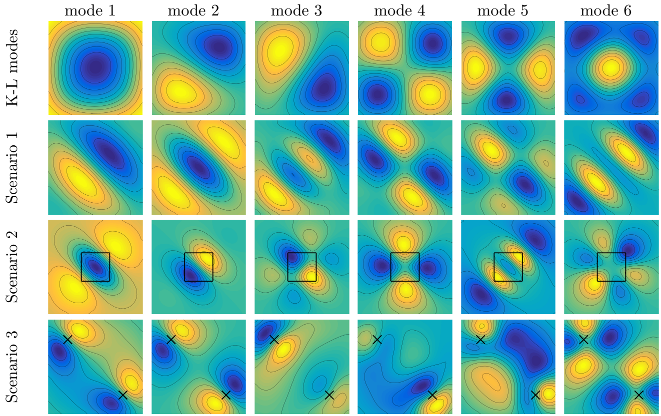
(b) log. diffusion field



(c) solution

1. Scenario 1 $\mathcal{M} : x \mapsto u \in V \subset H^1(\Omega)$
2. Scenario 2 $\mathcal{M} : x \mapsto u_{|\Omega_s} \in V \subset H^1(\Omega_s)$
3. Scenario 3 $\mathcal{M} : x \mapsto (u_{|s_1}, u_{|s_2}) \in V = \mathbb{R}^2$ (canonical norm)

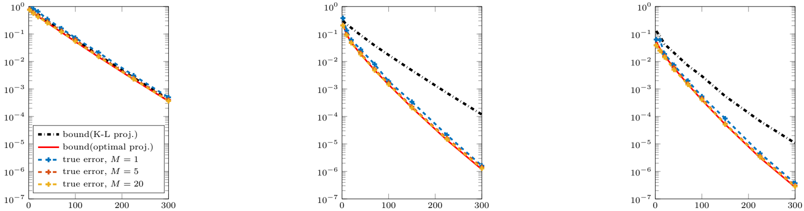
Modes $v_1, v_2, \ldots$



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

Approximation of the conditional expectation assuming H is known

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



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

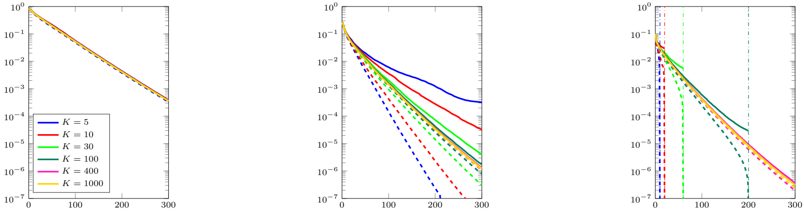
$$\|\mathcal{M} - \hat{F}_r\| = function(r)$$

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

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



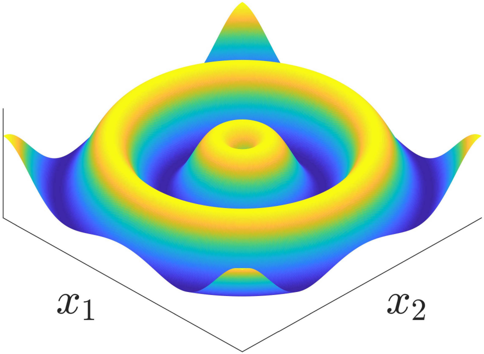
$\mathcal{M} : x \mapsto u$ $\qquad$ $\mathcal{M} : x \mapsto u_{|\Omega_s}$ $\qquad$ $\mathcal{M} : x \mapsto (u_{|s_1}, u_{|s_2})$

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

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

Notice that $\mathrm{rank}(\widehat{H}) \leq K \ \max_{1\leq k\leq K} \mathrm{rank}\left(\nabla \mathcal{M}(\mathbf{X}^{(k)})\right) \leq K \dim(V)$

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

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

A natural choice for J is then

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

Choice of the cost function J

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

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

A natural choice for J is then

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

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:

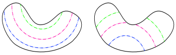
Assume $\mathcal{M} \in \mathcal{C}^1(\mathcal{X};\mathbb{R})$ and $\mathcal{G}_r \subset \mathcal{C}^1(\mathcal{X};\mathbb{R}^r)$.
Let $g : \mathcal{X} \to \mathbb{R}^r$ be a smooth function such that the level-sets

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

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?



yes! no

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

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

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

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

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

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

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

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

It is equivalent to seek for g solving

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

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

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, Migliorati, 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).

Once g is computed, how to construct f?

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

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):

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

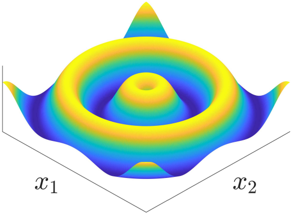
Illustration: isotropic function

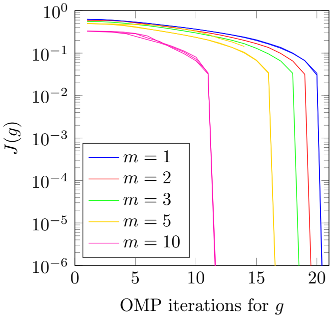
$$\mathcal{M}(\mathbf{x}) = \cos\left(\sqrt{x_1^2 + \ldots + x_d^2}\right)$$

$$\mu = \mathcal{N}(0, I_d)$$

$$\mathbf{x} \in \mathbb{R}^{20}$$

$$N = 100$$





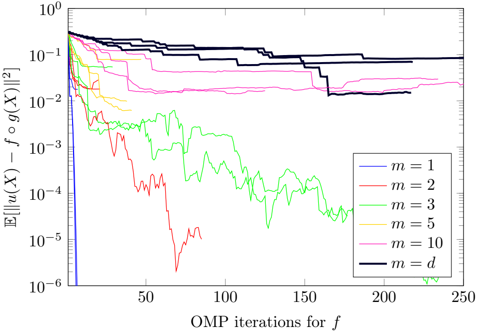
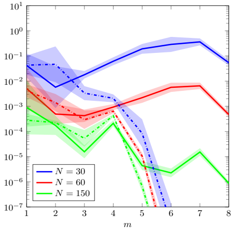
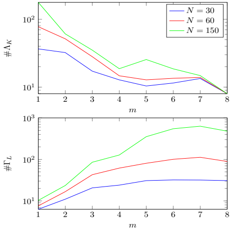


Illustration: Borehole function

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





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.

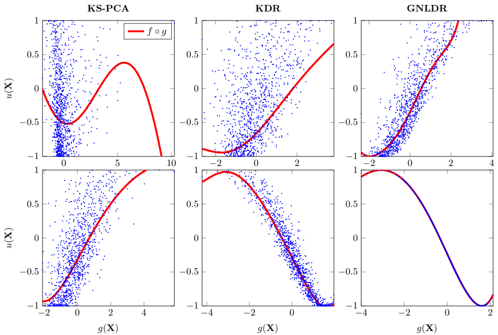
Comparison with nonlinear (NL) kernel supervised PCA and NL kernel dimension reduction.

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

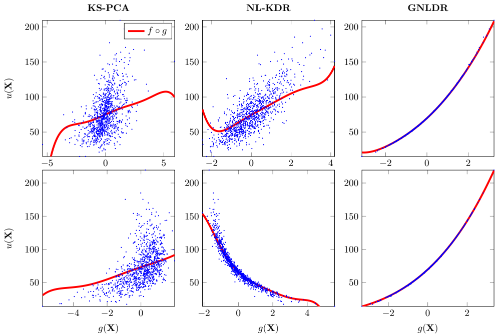
Kernel supervised PCA [Barshan et al., 2011] aims to maximize the dependence between $G^T\Phi(\mathbf{X})$ and $\mathbf{Y}$ measured with the Hilbert-Schmidt norm of the cross-covariance operator restricted to an arbitrary reproducing kernel Hilbert space (RKHS).

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.

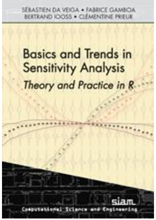
Perspectives

- It would be interesting to propose an optimal (or at least a clever) sampling procedure.
- 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.

Thanks for your attention!

Thanks for your attention!

And a little bit of advertisement



Some references I

Barshan, E., Ghodsi, A., Azimifar, Z., and Jahromi, M. Z. (2011).
Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds.
Pattern Recognition, 44(7):1357–1371.

Bigoni, D., Marzouk, Y., Prieur, C., and Zahm, O. (to appear, 2022).
Nonlinear dimension reduction for surrogate modeling using gradient information.
Information and Inference: A Journal of the IMA.

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

Da Veiga, S., Gamboa, F., Iooss, B., and Prieur, C. (2021).
Basics and Trends in Sensitivity Analysis: Theory and Practice in R.
SIAM.

Fukumizu, K., Bach, F. R., and Jordan, M. I. (2009).
Kernel dimension reduction in regression.
The Annals of Statistics, 37(4):1871–1905.

Some references II

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

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

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

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

Some references III

Zahm, O., Constantine, P. G., Prieur, C., and Marzouk, Y. M. (2020).
Gradient-based dimension reduction of multivariate vector-valued functions.
SIAM Journal on Scientific Computing, 42(1):A534–A558.