Global sensitivity analysis and dimension reduction

Clémentine PRIEUR
Grenoble Alpes University

Summer school on GSA & Poincaré inequalities
6-8 July 2022, Toulouse

Part II

Sensitivity analysis and dimension reduction:
a tour from sensitivity analysis to gradient based (non)linear dimension reduction via the so-called active subspace methodology.

Joint work with O. Zahm (Inria Grenoble), D. Bigoni & Y. Marzouk (MIT, Boston), P. Constantine (Univ. Colorado Boulder).

Notation

$$Y = \mathcal{M}(X_1, \ldots, X_d)$$

- $\mathbf{X} \sim \mu$: vector of input parameters with known probability distribution μ (not necessarily product measure)
- Y: output of interest, generally scalar $Y \in \mathbb{R}$
- $\mathcal{M}$: numerical model from it is also possible to evaluate the gradient

$$\nabla \mathcal{M}(\mathbf{x}) = \begin{pmatrix} \partial_1 \mathcal{M}(\mathbf{x}) \\ \vdots \\ \partial_d \mathcal{M}(\mathbf{x}) \end{pmatrix}$$

Variance based global sensitivity analysis: determine the relative influence of the inputs $X_1, \ldots, X_d$ on the output Y.

Gradient based GSA: how to best exploit the gradient $\nabla \mathcal{M}(\mathbf{x})$?

Previously...

Assume $\mu(d\mathbf{x}) = \prod_{i=1}^{d} \mu_i(dx_i)$ is a product measure. Then, for $i \in \{1, \ldots, d\}$, the total Sobol' indices

$$S_i^{\text{tot}} := 1 - \frac{\mathrm{Var}[\mathbb{E}(\mathcal{M}(\mathbf{X})|X_{-i})]}{\mathrm{Var}[\mathcal{M}(\mathbf{X})]}$$

can be bounded by

$$\mathrm{Var}[\mathcal{M}(\mathbf{X})]\, S_i^{\text{tot}} \leq C(\mu_i) \underbrace{\int (\partial_i \mathcal{M}(\mathbf{x}))^2 \mathrm{d}\mu(\mathbf{x})}_{=\nu_i}$$

where:

- $C(\mu_i)$ is the Poincaré constant of the marginal μ_i which is computable via an eigenvalue problem,
- ν_i is the i-th Derivative Based Sensitivity Measure (DGSM) which is often cheaper to compute compared to S_i^{tot}, provided $\partial_i \mathcal{M}$ is available.

1. Global sensitivity analysis from an approximation point of view

2. Gradient-based linear dimension reduction with active subspace

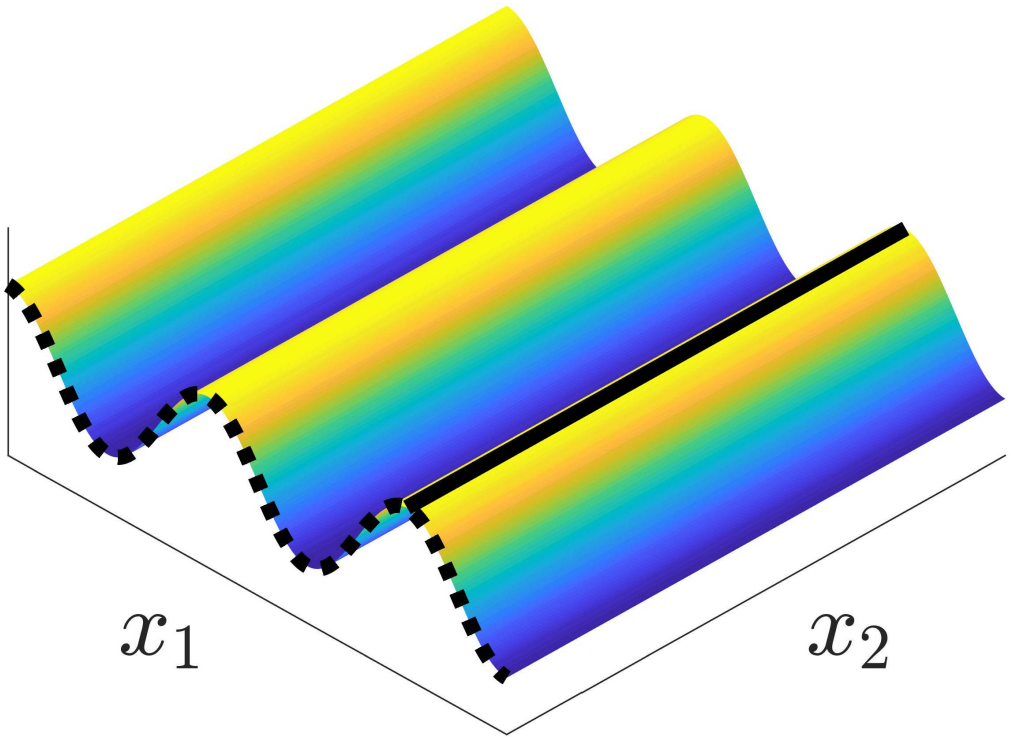
3. Extension to nonlinear dimension reduction

1. Global sensitivity analysis from an approximation point of view

2. Gradient-based linear dimension reduction with active subspace

3. Extension to nonlinear dimension reduction

We seek for a set of canonical coordinates along which the model is (almost) constant $\mathcal{M}(\mathbf{x}) \approx f(\mathbf{x_u})$



Hart and Gremaud [2018]

Function approximation perspective

Let $\mathbb{L}^2_\mu$ be the space of square-integrable functions u endowed with the norm

$$\|u\|^2 = \int u(\mathbf{x})^2 \mu(d\mathbf{x})\,.$$

Expectations and conditional expectations can be seen as $\mathbb{L}^2_\mu$ projections:

Function approximation perspective

Let $\mathbb{L}^2_\mu$ be the space of square-integrable functions u endowed with the norm

$$\|u\|^2 = \int u(\mathbf{x})^2 \mu(d\mathbf{x})\,.$$

Expectations and conditional expectations can be seen as $\mathbb{L}^2_\mu$ projections:

- the constant $c \in \mathbb{R}$ which best approximates $u(\mathbf{X})$ in $\mathbb{L}^2_\mu$ is $c = \mathbb{E}[u(\mathbf{X})]$:

$$\min_{c \in \mathbb{R}} \|u - c\|^2 = \|u - \mathbb{E}[u(\mathbf{X})]\|^2 =: \mathrm{Var}[u(\mathbf{X})];$$

Function approximation perspective

Let $\mathbb{L}^2_\mu$ be the space of square-integrable functions u endowed with the norm

$$\|u\|^2 = \int u(\mathbf{x})^2 \mu(d\mathbf{x})\,.$$

Expectations and conditional expectations can be seen as $\mathbb{L}^2_\mu$ projections:

- the constant $c \in \mathbb{R}$ which best approximates $u(\mathbf{X})$ in $\mathbb{L}^2_\mu$ is $c = \mathbb{E}[u(\mathbf{X})]$:

$$\min_{c \in \mathbb{R}} \|u - c\|^2 = \|u - \mathbb{E}[u(\mathbf{X})]\|^2 =: \mathrm{Var}[u(\mathbf{X})];$$

- the measurable function $f :\mapsto f(\mathbf{x_u}) = f(x_{u_1}, \ldots, x_{u_r})$ which best approximates $u(\mathbf{X})$ in $\mathbb{L}^2_\mu$ is $f(\mathbf{x_u}) = \mathbb{E}[u(\mathbf{X})|\mathbf{X_u} = \mathbf{x_u}]$:

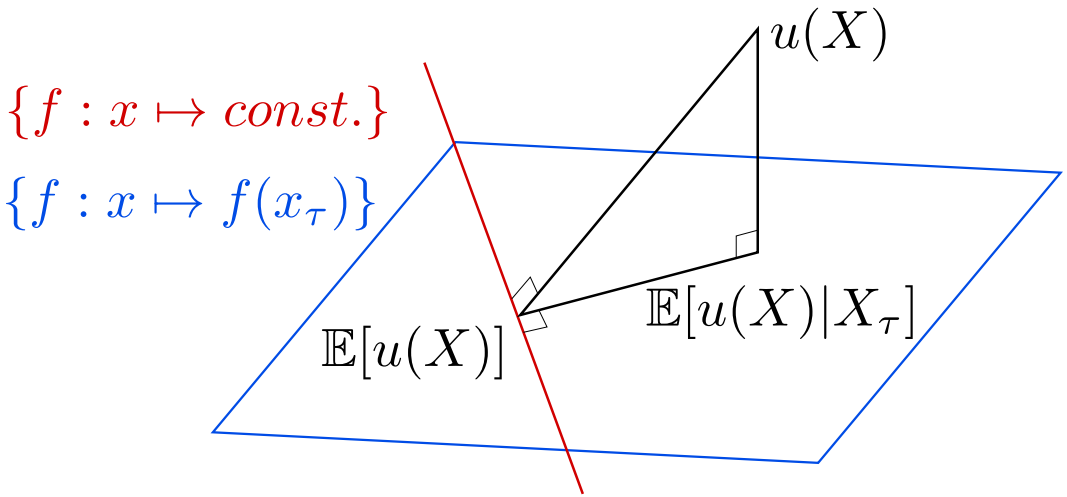
$$\min_{f:\mathbf{x}\mapsto f(\mathbf{x_u})\,,\ f \text{ measurable}} \|u - f\|^2 = \|u - \mathbb{E}[u(\mathbf{X})|\mathbf{X_u}]\|^2.$$

The total variance formula

$$\min_{f:\mathbf{x}\mapsto f(\mathbf{x_u})\,,\ f \text{ meas.}} \|u-f\|^2 = \|u-\mathbb{E}[u(\mathbf{X})|\mathbf{X_u}]\|^2$$

$$= \|\Big(u-\mathbb{E}[u(\mathbf{X})]\Big) - \Big(\mathbb{E}[u(\mathbf{X})]-\mathbb{E}[u(\mathbf{X})|\mathbf{X_u}]\Big)\|^2$$

$$= \underbrace{\|u-\mathbb{E}[u(\mathbf{X})]\|^2}_{\mathrm{Var}[u(\mathbf{X})]} - \underbrace{\|\mathbb{E}[u(\mathbf{X})]-\mathbb{E}[u(\mathbf{X})|\mathbf{X_u}]\|^2}_{\mathrm{Var}[\mathbb{E}(u(\mathbf{X})|\mathbf{X_u})]} \ (\star)$$



Link with Sobol' indices

The closed Sobol' indices for $\mathcal{M}$ writes

$$S_{\mathbf{u}}^{\text{clo}} := \frac{\mathrm{Var}[\mathbb{E}(\mathcal{M}(\mathbf{X})|\mathbf{X}_{\mathbf{u}})]}{\mathrm{Var}[\mathcal{M}(\mathbf{X})]} \overset{(\star)}{=} 1 - \frac{\min\limits_{f:\mathbf{x}\mapsto f(\mathbf{x}_{\mathbf{u}})\,,\, f \text{ meas.}} \|\mathcal{M} - f\|^2}{\mathrm{Var}[\mathcal{M}(\mathbf{X})]}$$

$$\boxed{\begin{aligned} S_{\mathbf{u}}^{\text{clo}} \approx 1 \quad &\Leftrightarrow \quad \mathcal{M}(\mathbf{X}) \approx f(\mathbf{X}_{\mathbf{u}}) \\ &\Leftrightarrow \quad \mathbf{X}_{\mathbf{u}} \text{ “explains” well } Y = \mathcal{M}(\mathbf{X}) \end{aligned}}$$

Define the total Sobol' index for $\mathcal{M}$ associated to subset $\mathbf{u} \subseteq \{1, \ldots, d\}$ as

$$S_{\mathbf{u}}^{\text{tot}} = \frac{\mathbb{E}[\mathrm{Var}(\mathcal{M}(\mathbf{X})|\mathbf{X}_{-\mathbf{u}})]}{\mathrm{Var}(\mathcal{M}[\mathbf{X})]}.$$

We have

$$S_{\mathbf{u}}^{\text{tot}} \overset{(\star)}{=} \frac{\min\limits_{f:\mathbf{x}\mapsto f(\mathbf{x}_{-\mathbf{u}})\,,\, f \text{ meas.}} \|\mathcal{M} - f\|^2}{\mathrm{Var}[\mathcal{M}(\mathbf{X})]}$$

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

Define the total Sobol' index for $\mathcal{M}$ associated to subset $\mathbf{u} \subseteq \{1, \ldots, d\}$ as

$$S_{\mathbf{u}}^{\text{tot}} = \frac{\mathbb{E}[\mathrm{Var}(\mathcal{M}(\mathbf{X})|\mathbf{X}_{-\mathbf{u}})]}{\mathrm{Var}(\mathcal{M}[\mathbf{X})]}.$$

We have

$$S_{\mathbf{u}}^{\text{tot}} \overset{(\star)}{=} \frac{\min\limits_{f:\mathbf{x}\mapsto f(\mathbf{x}_{-\mathbf{u}})\,,\, f \text{ meas.}} \|\mathcal{M} - f\|^2}{\mathrm{Var}[\mathcal{M}(\mathbf{X})]}$$

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

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

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

In that framework, if $S_{\mathbf{u}}^{\text{tot}} \approx 0$, we get:

$$\mathcal{M}(\mathbf{X}_{\mathbf{u}} : \mathbf{X}_{-\mathbf{u}}) \approx \mathcal{M}(\mathbf{x}_{\mathbf{u}}^{0} : \mathbf{X}_{-\mathbf{u}})\,.$$

1. Global sensitivity analysis from an approximation point of view

2. Gradient-based linear dimension reduction with active subspace

3. Extension to nonlinear dimension reduction

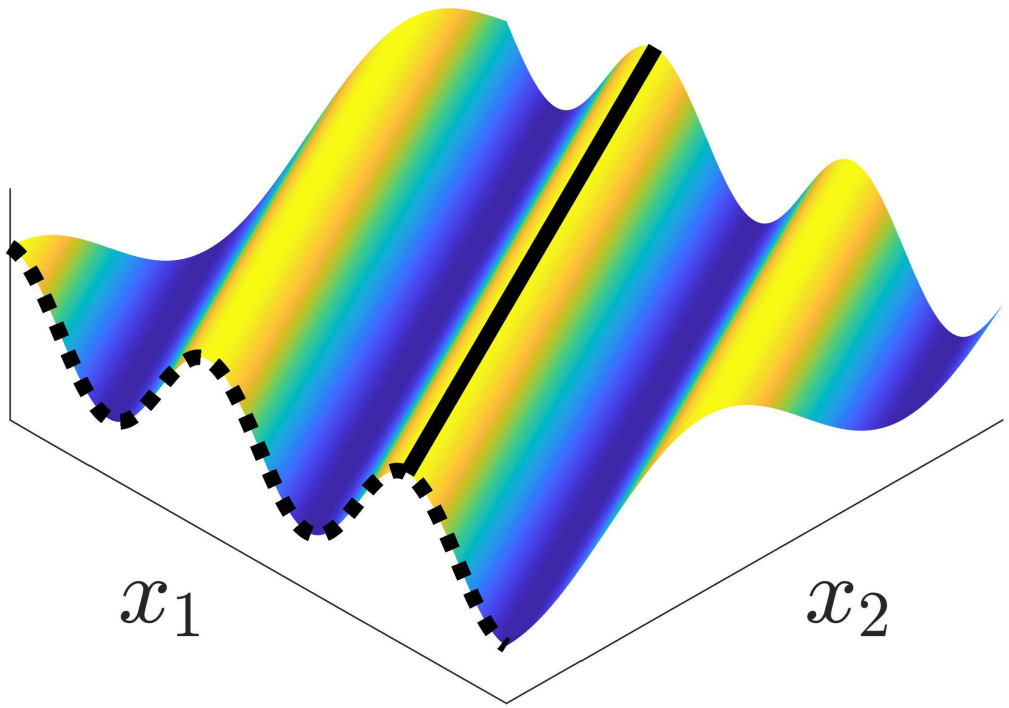
Gradient based linear dimension reduction

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

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 $r \leq d$.



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 $\| \cdot \|_{\mathcal{Y}}$.

Let $\mathbb{L}^2_\mu$ be the space of square-integrable functions $u : \mathbb{R}^d \to \mathcal{Y}$. We endow it with the norm

$$\|u\|^2 = \int \|u(\mathbf{x})\|_{\mathcal{Y}}{}^2 \mu(d\mathbf{x}) = \mathbb{E}\left[\|u(\mathbf{X})\|_{\mathcal{Y}}{}^2\right] .$$

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 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,

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

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}):

$$H v_i = \lambda_i \Sigma^{-1} v_i \,, \;\; {v_i}^T \Sigma^{-1} v_j = \delta_{i,j}.$$

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

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

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

- 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.

Basic examples with $p = 1$

- Affine function $\mathcal{M}(\mathbf{x}) = \alpha^T \mathbf{x} + \beta$. Then $\nabla \mathcal{M}(\mathbf{x}) = \alpha$ and $H = \alpha\alpha^T$.

Basic examples with $p = 1$

- Affine function $\mathcal{M}(\mathbf{x}) = \alpha^T \mathbf{x} + \beta$. Then $\nabla \mathcal{M}(\mathbf{x}) = \alpha$ and $H = \alpha\alpha^T$.

- Ridge function $\mathcal{M}(\mathbf{x}) = f(\alpha^T \mathbf{x})$ for some $\alpha \in \mathbb{R}^d$. We have $\nabla \mathcal{M}(\mathbf{x}) = \alpha f'(\alpha^T \mathbf{x})$ and $H = \alpha^T \left(\int f'(\alpha^T \mathbf{x})^2 \mathrm{d}\mu(\mathbf{x}) \right) \alpha$.

Basic examples with $p = 1$

- Affine function $\mathcal{M}(\mathbf{x}) = \alpha^T \mathbf{x} + \beta$. Then $\nabla \mathcal{M}(\mathbf{x}) = \alpha$ and $H = \alpha\alpha^T$.

- Ridge function $\mathcal{M}(\mathbf{x}) = f(\alpha^T \mathbf{x})$ for some $\alpha \in \mathbb{R}^d$. We have $\nabla \mathcal{M}(\mathbf{x}) = \alpha f'(\alpha^T \mathbf{x})$ and $H = \alpha^T \left(\int f'(\alpha^T \mathbf{x})^2 \mathrm{d}\mu(\mathbf{x}) \right) \alpha$.

- Generalized ridge function $\mathcal{M}(\mathbf{x}) = f(V_r^T \mathbf{x})$, $V_r \in \mathbb{R}^{d \times r}$. Then $H = V_r \left(\int \nabla f(V_r^T \mathbf{x}) \nabla f(V_r^T \mathbf{x})^T \mathrm{d}\mu(\mathbf{x}) \right) V_r^T$.

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

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

See, e.g., Sobol' & Kucherenko, 2009 and Lamboni *et al.*, 2013 for similar results in the case $p = 1$ (scalar output) and $\mathbf{u}$ a singleton. Sobol' indices for vectorial outputs have been introduced in Lamboni et al. [2011] and further studied in Gamboa et al. [2013].

Also, 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:

$$S_{\mathbf{u}}^{\text{clo}} = 1 - \frac{\|\mathcal{M} - \mathbb{E}_\mu(\mathcal{M}|\sigma(P_r))\|_{\mathcal{H}}^2}{\|\mathcal{M} - \mathbb{E}_\mu(\mathcal{M})\|_{\mathcal{H}}^2}$$

thus

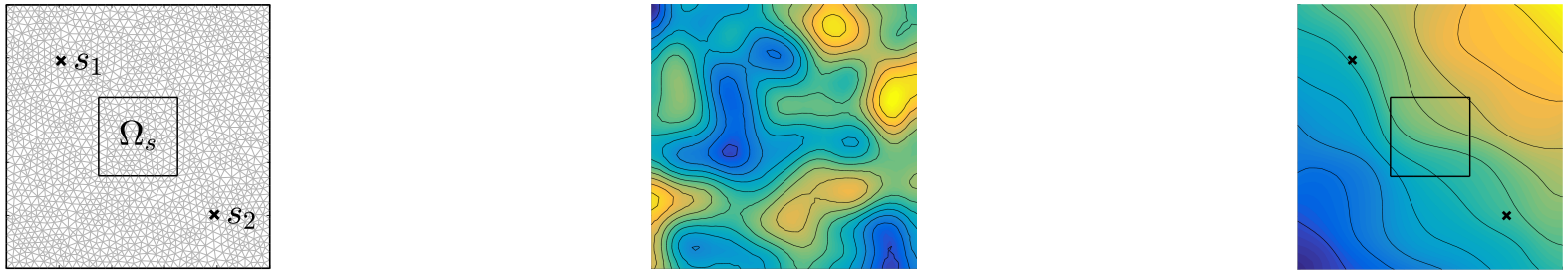
$$\begin{aligned} S_{\mathbf{u}}^{\text{clo}} \quad & \geq \quad 1 - \frac{\text{trace}\left(\Sigma P_r^T H P_r\right)}{\|\mathcal{M} - \mathbb{E}_\mu(\mathcal{M})\|^2} \\ & \underbrace{=}_{\text{for independent inputs}} \quad 1 - \frac{\sum_{i \in -\mathbf{u}} \mathrm{Var}(X_i) H_{i,i}}{\|\mathcal{M} - \mathbb{E}_\mu(\mathcal{M})\|^2} \ . \end{aligned}$$

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:

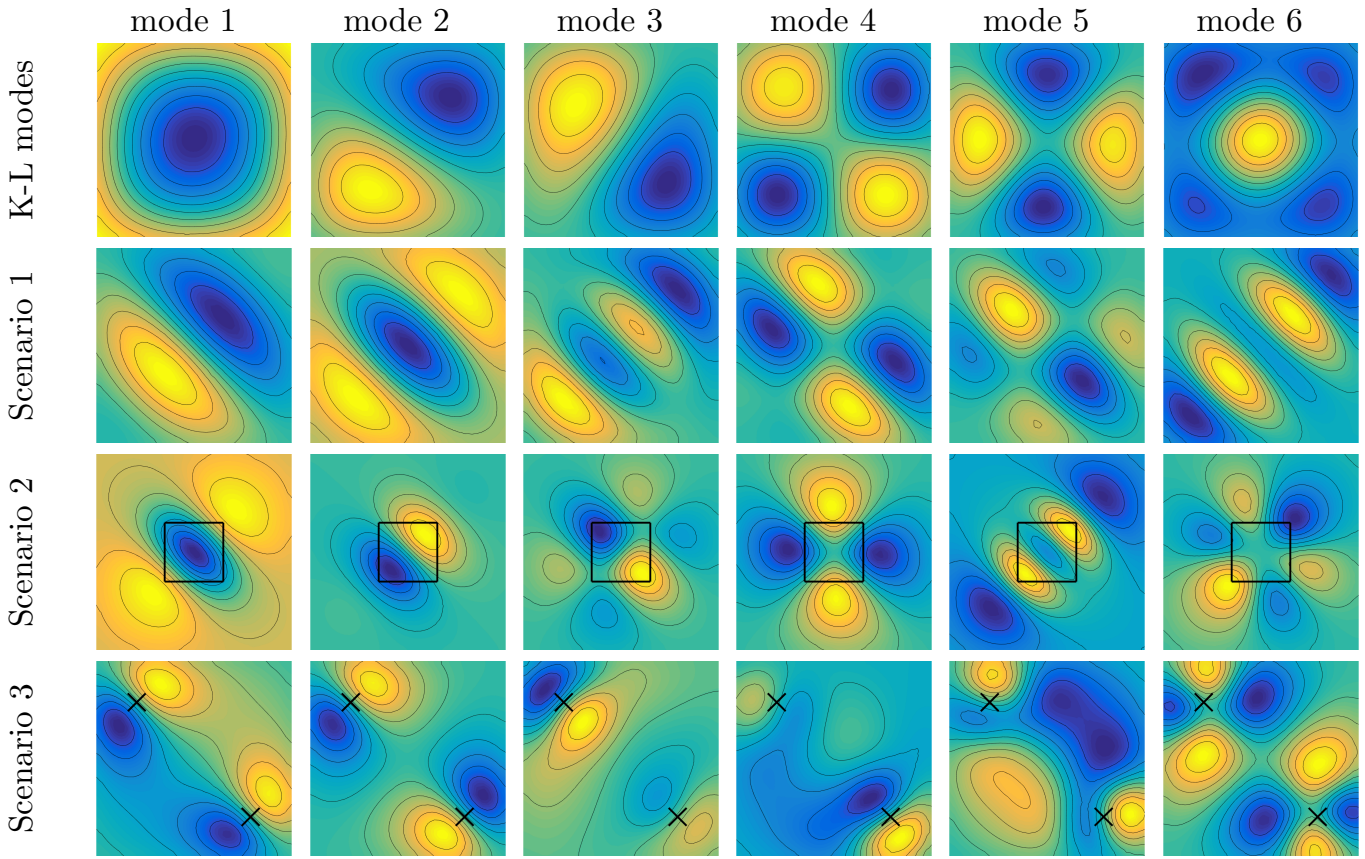
$$x = \log(\kappa) \in \mathbb{R}^{3252} \sim \mu = \mathcal{N}(0, \Sigma)$$



(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).

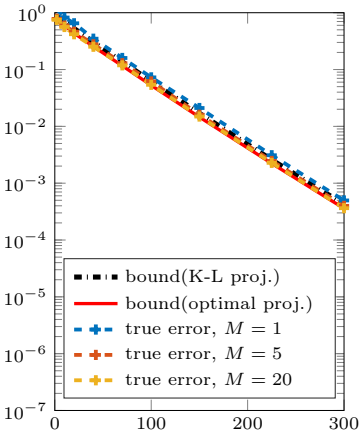
Modes $v_1, v_2, \ldots$



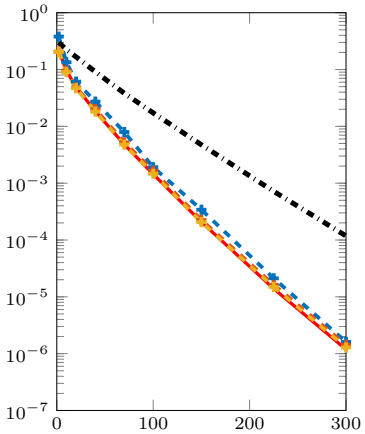
$$\mathrm{Im}(P_r) = \mathrm{span}\{v_1, v_2, \ldots, v_r\}$$

Approximation of the conditional expectation assuming H is known

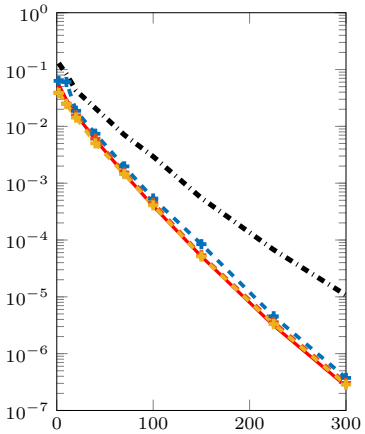
$$\mathbb{E}_\mu(\mathcal{M}|\sigma(P_r)) \approx \hat{F}_r : x \mapsto \frac{1}{M}\sum_{k=1}^{M} \mathcal{M}(P_r x + (I_d - P_r)\mathbf{Z}^{(k)}), \quad \mathbf{Z}^{(k)} \overset{iid}{\sim} \mu$$



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



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



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

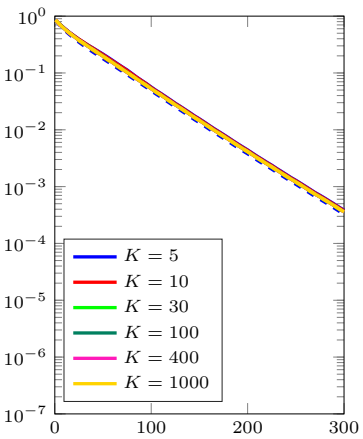
$$\|\mathcal{M} - \hat{F}_r\| = function(r)$$

We can show that

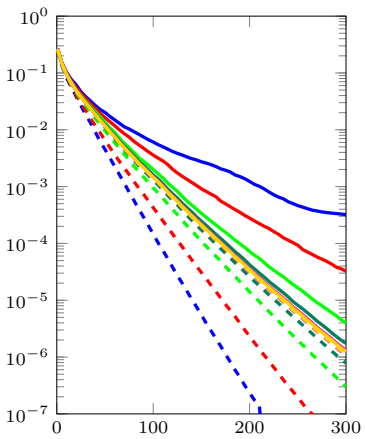
$$\mathbb{E}\left(\|\mathcal{M} - \hat{F}_r\|^2\right) \leq (1 + M^{-1})\ \mathrm{trace}(\Sigma(I_d - P_r^T)H(I_d - P_r))$$

Approximation of H to get the projector

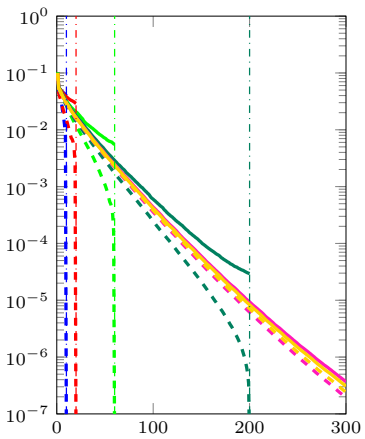
$$H \approx \widehat{H} = \frac{1}{K}\sum_{k=1}^{K}(\nabla\mathcal{M}(\mathbf{X}^{(k)}))^*(\nabla\mathcal{M}(\mathbf{X}^{(k)})), \quad \mathbf{X}^{(k)} \overset{iid}{\sim} \mu$$



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



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



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

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

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

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

(see also Zahm et al. [2022])

Beyond Gaussian uncertainty

Let $\mathrm{d}\mu(x) \propto \exp\Big(-V(x) - \Psi(x)\Big)\mathrm{d}x$. 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]):

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

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

Beyond Gaussian uncertainty

Let $\mathrm{d}\mu(x) \propto \exp\Big(-V(x) - \Psi(x)\Big)\mathrm{d}x$. 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]):

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

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

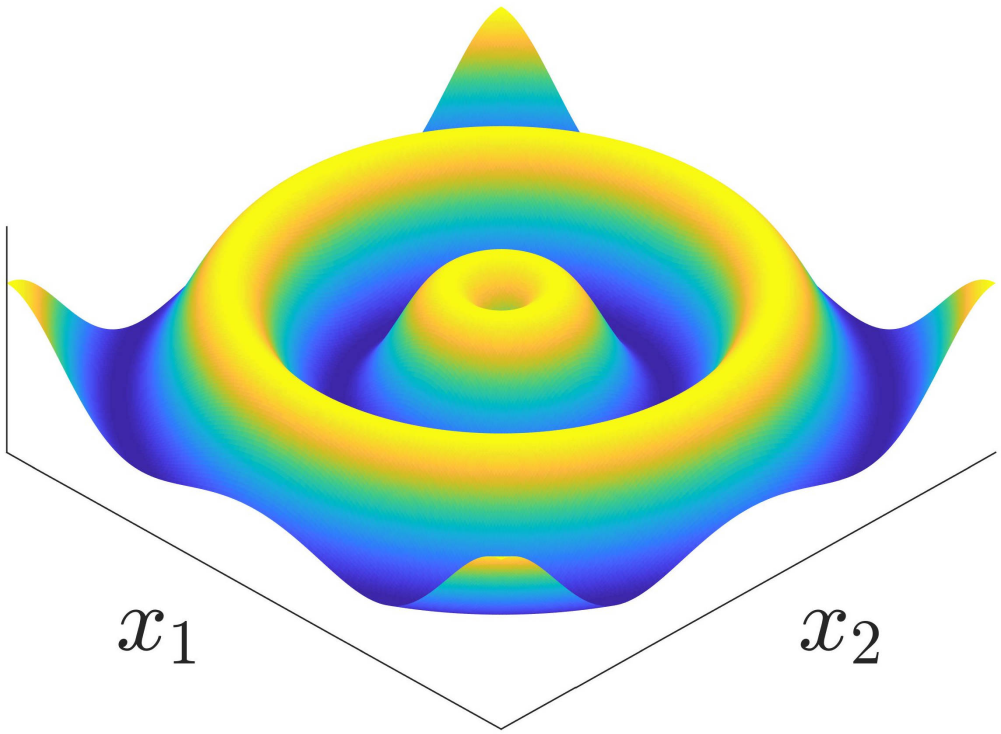
- Gaussian mixtures,
- uniform measures on compact & convex sets
- any measure such that $\mathrm{d}\mu(\mathbf{x}) \geq \alpha > 0$ on compact & convex sets.

1. Global sensitivity analysis from an approximation point of view

2. Gradient-based linear dimension reduction with active subspace

3. Extension to nonlinear dimension reduction

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



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

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

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

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

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:

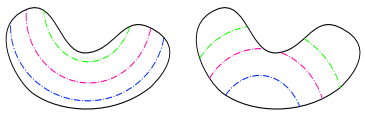
Assume $\mathcal{M} \in \mathcal{C}^1(\mathcal{X}; \mathbb{R})$ and $\mathcal{G}_r \subset \mathcal{C}^1(\mathcal{X}; \mathbb{R}^r)$.
Let $g : \mathcal{X} \to \mathbb{R}^r$ be a smooth function such that the level-sets

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

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

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.

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

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

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

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

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

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

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

It is equivalent to seek for g solving

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

$H(\mathbf{x}) = \nabla\Phi(\mathbf{x})(\nabla\mathcal{M}(\mathbf{x})\nabla\mathcal{M}(\mathbf{x})^T)\nabla\Phi(\mathbf{x})^T$,
$\Sigma(\mathbf{x}) = \nabla\Phi(\mathbf{x})\nabla\Phi(\mathbf{x})^T$, 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:

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

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).

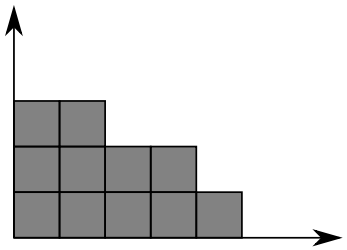
More precisely, to adapt the complexity of $\mathcal{G}$ with respect to the sample size N, one uses the following tools:

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:



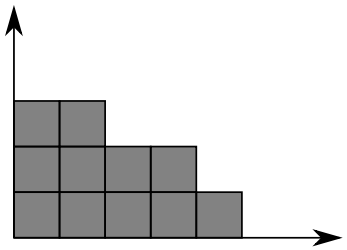
$$\nu \in \Lambda \text{ and } \mu \leq \nu \quad \Rightarrow \quad \mu \in \Lambda$$

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:

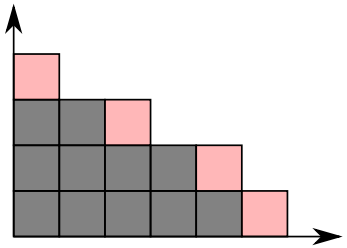


$$\nu \in \Lambda \text{ and } \mu \leq \nu \quad \Rightarrow \quad \mu \in \Lambda$$

Matching Pursuit

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

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



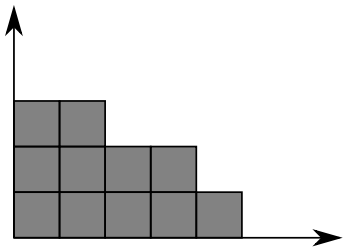
where G_k^* is the minimizer of $\hat{\mathcal{R}}^N(\cdot)$ over Λ_k.

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:

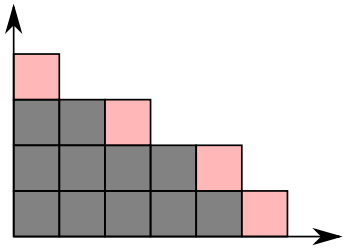


$$\nu \in \Lambda \text{ and } \mu \leq \nu \quad \Rightarrow \quad \mu \in \Lambda$$

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?

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

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?

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

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.

Benchmark algorithm (without dimension reduction):

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

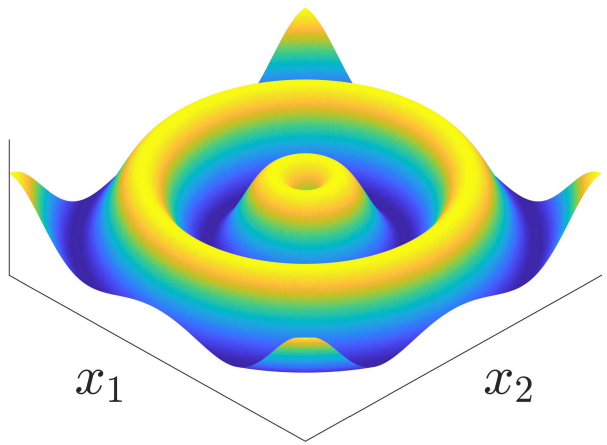
Illustration: isotropic function

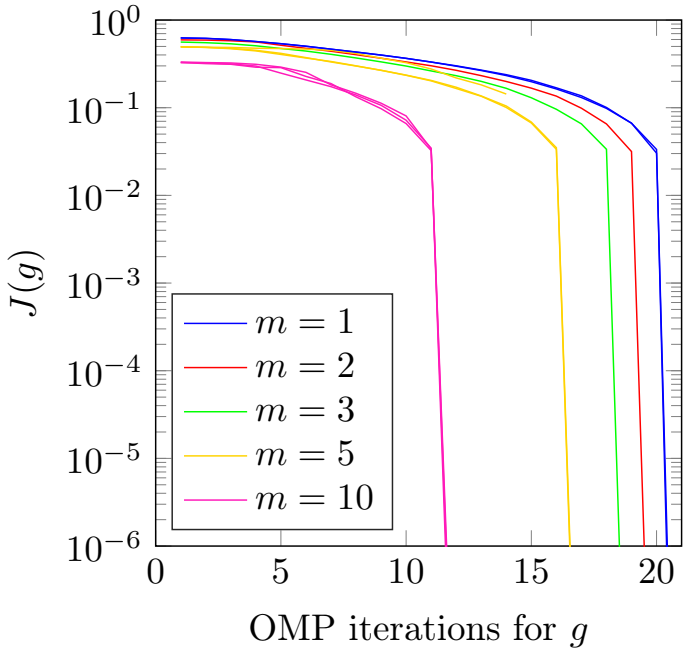
$$\mathcal{M}(\mathbf{x}) = \cos\left(\sqrt{x_1^2 + \ldots + x_d^2}\right)$$

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

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

$$N = 100$$





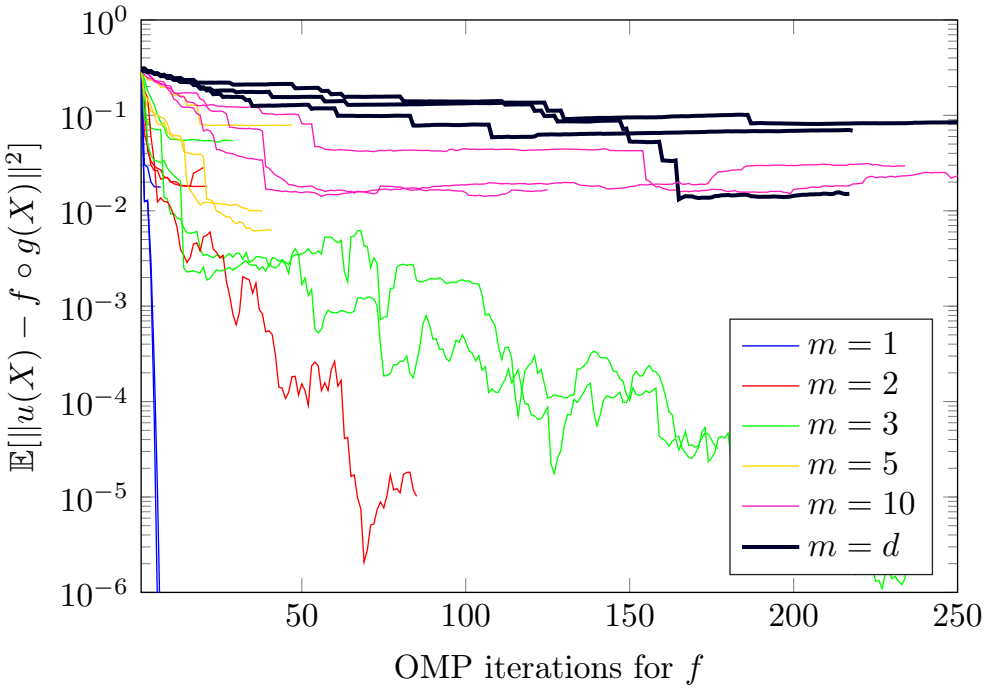
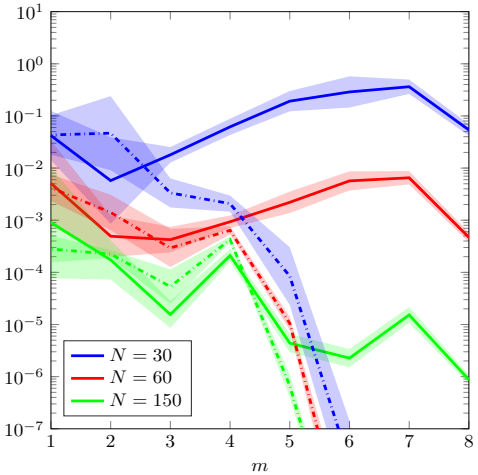
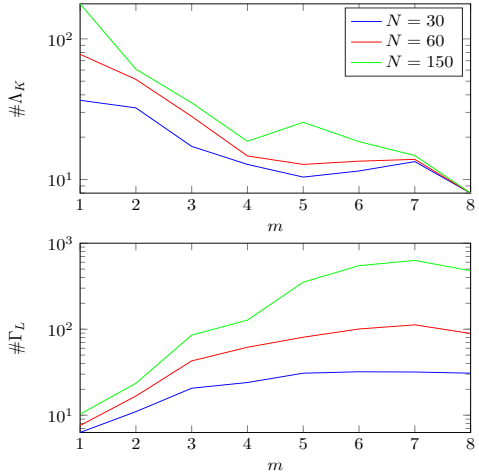


Illustration: Borehole function

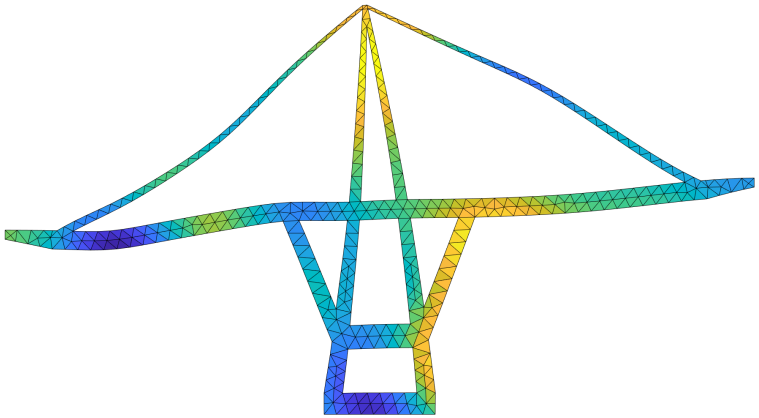
$$\mathcal{M}(\mathbf{x}) = \frac{2\pi T_{\mathcal{M}}(H_{\mathcal{M}} - H_\ell)}{\ln(r/r_\omega)\left(1 + \frac{2LT_{\mathcal{M}}}{\ln(r/r_\omega) r_\omega^2 K_\omega} + \frac{T_{\mathcal{M}}}{T_\ell}\right)}, \quad \begin{cases} x_1 = r_\omega & \sim \mathcal{N}(0.1, 3 \cdot 10^{-4}) \\ x_2 = r & \sim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





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



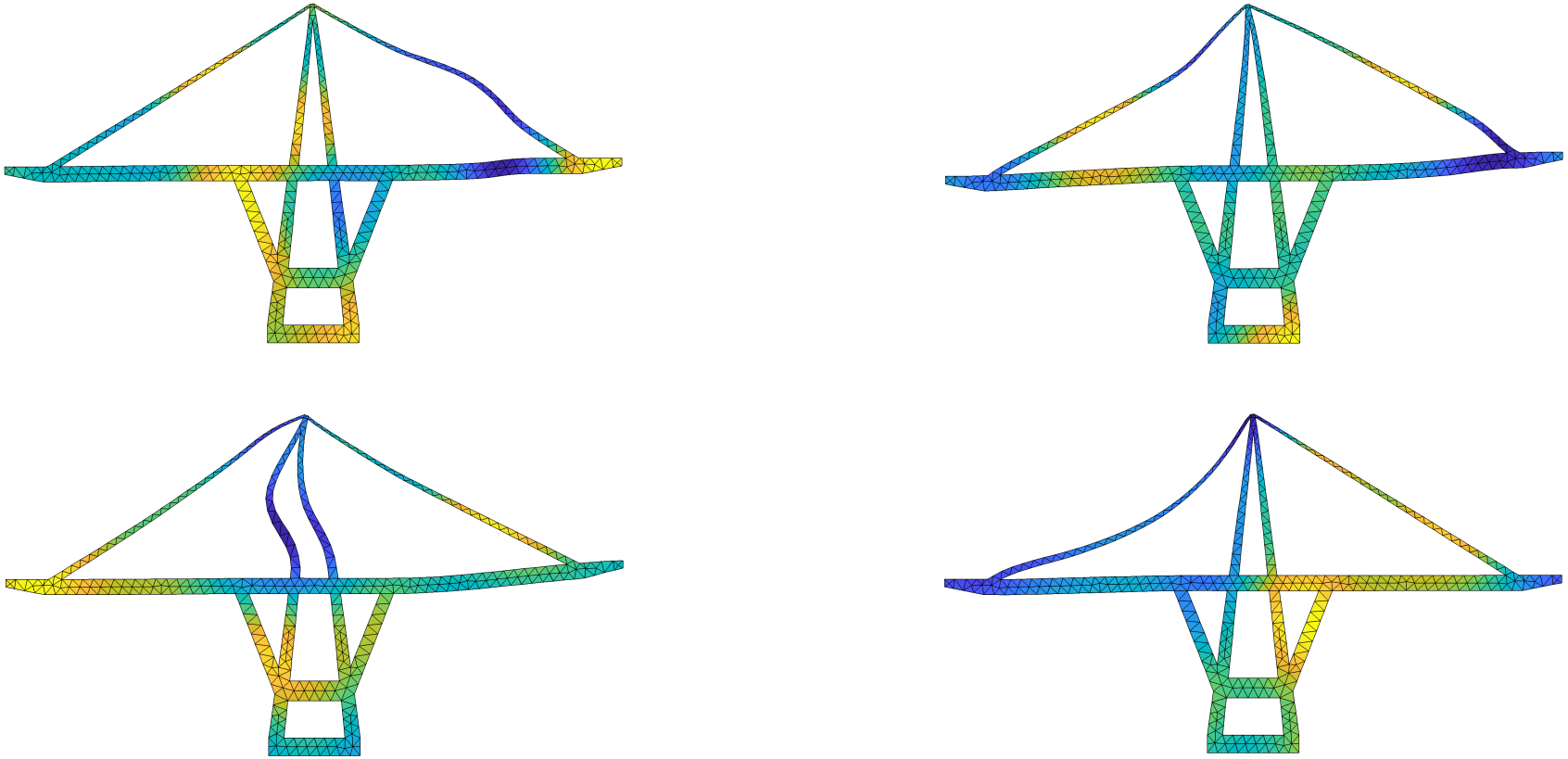
Parametrized eigenvalue problem

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

- $K(\mathbf{x})$: stiffness matrix
- M: mass matrix
- $v \in \mathbb{R}^{\mathcal{N}}$: Finite Element solution ($\mathcal{N} = 960$)
- $\mathbf{x} \in \mathbb{R}^d$: 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 $\mathbf{X}$ (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 $\mathbb{E}[(\mathcal{M}(\mathbf{X}) - f \circ g(\mathbf{X}))^2]$ 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.

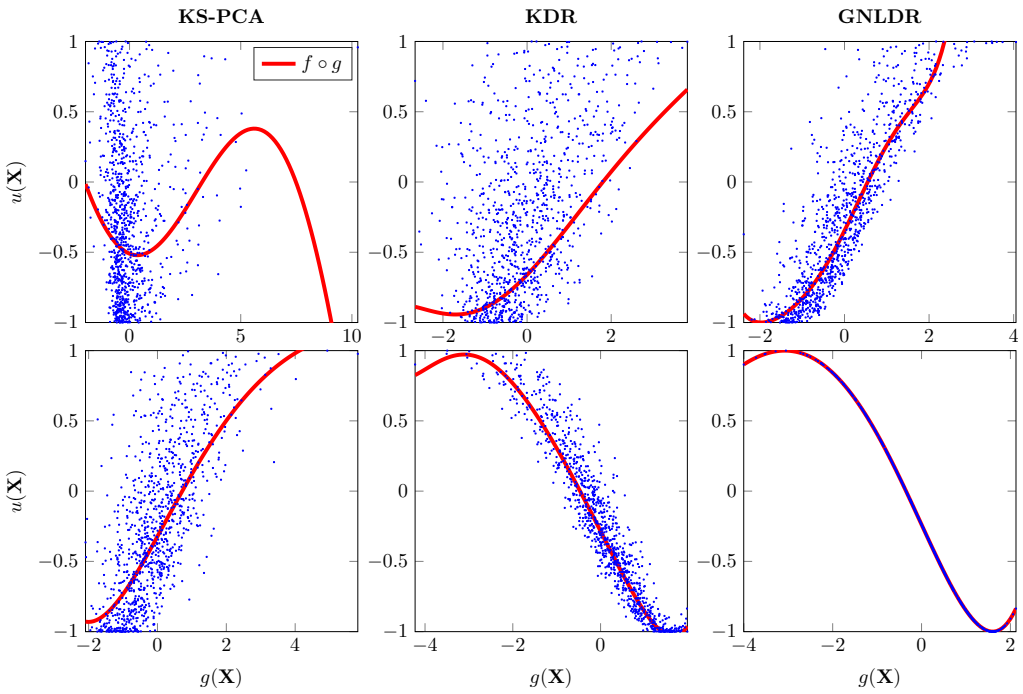
Comparison with nonlinear (NL) kernel supervised PCA and NL kernel dimension reduction.

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

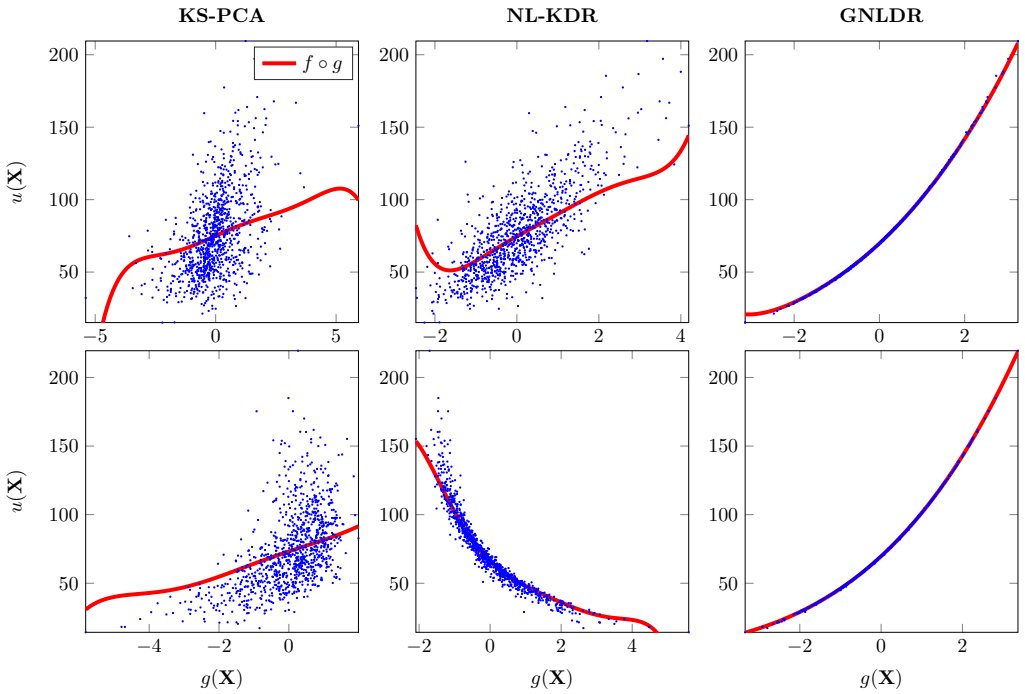
Kernel supervised PCA Barshan et al. [2011] aims to maximize the dependence between $G^T \Phi(\mathbf{X})$ and $\mathbf{Y}$ measured with the Hilbert-Schmidt norm of the cross-covariance operator restricted to an arbitrary reproducing kernel Hilbert space (RKHS).

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

- During this lecture, 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 algorithm takes into account gradient information.

Conclusion

- During this lecture, 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 algorithm takes into account gradient information.

Perspectives

- It would be interesting to propose an optimal (or at least a clever) sampling procedure.
- Although assuming $\mathbb{C}(\mathbf{X}|\mathcal{G}_r) < \infty$ is usual, proving it remains an open challenge. A solution to investigate would be to construct g in a way that ensures $\mathbb{P}_{\mathbf{X}|\mathcal{G}_r}$ to be the push-forward measure of the standard normal distribution through a Lipschitz map.

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*, 2022. 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.

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`.

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.

Some references II

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.

Olivier Zahm, Tiangang Cui, Kody Law, Alessio Spantini, and Youssef Marzouk. Certified dimension reduction in nonlinear bayesian inverse problems. *Mathematics of Computation*, 2022. doi: https://doi.org/10.1090/mcom/3737.

Thanks for your attention!