Structured Nonlinear Dimension Reduction Using Gradient Evaluations

Alexandre PASCO

Centrale Nantes, Nantes Université, France
In collaboration with Anthony NOUY.

Introduction

- Goal : Approximate $u : \mathbb{R}^d \to \mathbb{R} \in \mathcal{C}^1$ with $d \gg 1$, i.e. minimize

$$\mathcal{E}(\hat{u}) := \mathbb{E}\left[(u(X) - \hat{u}(X))^2\right]$$

where X has probability density μ_X.

- Given : Few costly point evaluations drawn from $\mu_X dx$,

$$\left(x^{(i)}, u(x^{(i)}), \nabla u(x^{(i)})\right)_{1 \le i \le n_s}.$$

- Remark : vector-valued u described in [ZCPM20, VPZ23]

- Goal : Approximate $u : \mathbb{R}^d \to \mathbb{R} \in \mathcal{C}^1$ with $d \gg 1$, i.e. minimize

$$\mathcal{E}(\hat{u}) := \mathbb{E}\left[(u(X) - \hat{u}(X))^2\right]$$

where X has probability density μ_X.

- Given : Few costly point evaluations drawn from $\mu_X dx$,

$$\left(x^{(i)}, u(x^{(i)}), \nabla u(x^{(i)})\right)_{1 \leq i \leq n_s}.$$

- Remark : vector-valued u described in [ZCPM20, VPZ23]

Introduction

- Goal : Approximate $u : \mathbb{R}^d \to \mathbb{R} \in \mathcal{C}^1$ with $d \gg 1$, i.e. minimize

$$\mathcal{E}(\hat{u}) := \mathbb{E}\left[(u(X) - \hat{u}(X))^2\right]$$

where X has probability density μ_X.

- Given : Few costly point evaluations drawn from $\mu_X dx$,

$$\left(x^{(i)}, u(x^{(i)}), \nabla u(x^{(i)})\right)_{1 \leq i \leq n_s}.$$

- Remark : vector-valued u described in [ZCPM20, VPZ23]

A first approach

- Gradient enhanced regression,

$$\min_{v \in \mathcal{U}_N} \frac{1}{n} \sum_{i=1}^{n} (u(x^{(i)}) - v(x^{(i)}))^2 + \|\nabla u(x^{(i)}) - \nabla v(x^{(i)})\|^2,$$

  taking $\mathcal{U}_N$ as polynomials, splines, Neural Nets, ...

- Good idea : improvement over classical least-squares.

- Problem : large d and few samples $\Rightarrow$ ill-conditioned problem.

A first approach

- Gradient enhanced regression,

$$\min_{v \in \mathcal{U}_N} \frac{1}{n} \sum_{i=1}^{n} (u(x^{(i)}) - v(x^{(i)}))^2 + \|\nabla u(x^{(i)}) - \nabla v(x^{(i)})\|^2,$$

taking $\mathcal{U}_N$ as polynomials, splines, Neural Nets, ...

- Good idea : improvement over classical least-squares.

- Problem : large d and few samples $\Rightarrow$ ill-conditioned problem.

A first approach

- Gradient enhanced regression,

$$\min_{v \in \mathcal{U}_N} \frac{1}{n} \sum_{i=1}^{n} (u(x^{(i)}) - v(x^{(i)}))^2 + \|\nabla u(x^{(i)}) - \nabla v(x^{(i)})\|^2,$$

  taking $\mathcal{U}_N$ as polynomials, splines, Neural Nets, ...

- Good idea : improvement over classical least-squares.

- Problem : large d and few samples $\Rightarrow$ ill-conditioned problem.

Our approach

- Approximation of the form $\tilde{u} = f \circ g$.
- Example : if u is a QoI of some PDE discretized with d dofs, then Reduced Basis means taking $g(x) = U_r^T x$.
- Step 1 : Learn a **feature map** $g \in \mathcal{G}_m \subseteq \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^m)$ with $m \leq d$, for some chosen **tractable** function class $\mathcal{G}_m$.
- Step 2 : Learn a **profile map** $f : \mathbb{R}^m \to \mathbb{R}$.

Our approach

- Approximation of the form $\tilde{u} = f \circ g$.
- Example : if u is a QoI of some PDE discretized with d dofs, then Reduced Basis means taking $g(x) = U_r^T x$.
- Step 1 : Learn a **feature map** $g \in \mathcal{G}_m \subseteq \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^m)$ with $m \leq d$, for some chosen **tractable** function class $\mathcal{G}_m$.
- Step 2 : Learn a **profile map** $f : \mathbb{R}^m \to \mathbb{R}$.

Our approach

- Approximation of the form $\tilde{u} = f \circ g$.
- Example : if u is a QoI of some PDE discretized with d dofs, then Reduced Basis means taking $g(x) = U_r^T x$.
- Step 1 : Learn a **feature map** $g \in \mathcal{G}_m \subseteq \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^m)$ with $m \leq d$, for some chosen **tractable** function class $\mathcal{G}_m$.
- Step 2 : Learn a **profile map** $f : \mathbb{R}^m \to \mathbb{R}$.

Our approach

- Approximation of the form $\tilde{u} = f \circ g$.
- Example : if u is a QoI of some PDE discretized with d dofs, then Reduced Basis means taking $g(x) = U_r^T x$.
- Step 1 : Learn a **feature map** $g \in \mathcal{G}_m \subseteq \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^m)$ with $m \leq d$, for some chosen **tractable** function class $\mathcal{G}_m$.
- Step 2 : Learn a **profile map** $f : \mathbb{R}^m \to \mathbb{R}$.

Our approach

- Approximation of the form $\tilde{u} = f \circ g$.
- Example : if u is a QoI of some PDE discretized with d dofs, then Reduced Basis means taking $g(x) = U_r^T x$.
- Step 1 : Learn a **feature map** $g \in \mathcal{G}_m \subseteq \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^m)$ with $m \leq d$, for some chosen **tractable** function class $\mathcal{G}_m$.
- Step 2 : Learn a **profile map** $f : \mathbb{R}^m \to \mathbb{R}$.
- Step 2 : Parametric studies in $\mathbb{R}^m$ instead of $\mathbb{R}^d$.

Our approach

- Approximation of the form $\tilde{u} = f \circ g$.
- Example : if u is a QoI of some PDE discretized with d dofs, then Reduced Basis means taking $g(x) = U_r^T x$.
- Step 1 : Learn a **feature map** $g \in \mathcal{G}_m \subseteq \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^m)$ with $m \leq d$, for some chosen **tractable** function class $\mathcal{G}_m$.
- Step 2 : Learn a **profile map** $f : \mathbb{R}^m \to \mathbb{R}$.

Profile map

- For a given g, the best profile function is

$$f_g(z) := \mathbb{E}\left[u(X)|Z = z\right],$$

where $Z := g(X) \in \mathbb{R}^m$. Problem : **not computable**.

- In practice : learn f via gradient enhanced regression, but with only $m \leq d$ input variables.

Profile map

- For a given g, the best profile function is

$$f_g(z) := \mathbb{E}\left[u(X)|Z = z\right],$$

where $Z := g(X) \in \mathbb{R}^m$. Problem : **not computable**.

- In practice : learn f via gradient enhanced regression, but with only $m \leq d$ input variables.

Measuring the quality of g

What if $u = f \circ g$?

for some $g \in \mathcal{G}_m$ and some f.

If $u = f \circ g$

- Assume $u = f \circ g$ then $\nabla u(x) = J_g^T(x) J_f(g(x))$.Thus

$$\exists f \text{ s.t. } u = f \circ g \quad \Longrightarrow \quad \nabla u(x) \in \text{cspan}(J_g^T(x)), \forall x$$

- If $g^{-1}(z)$ is smoothly pathwise-connected for all $z \in \text{Im}(g)$,

$$\exists f \text{ s.t. } u = f \circ g \quad \Longleftrightarrow \quad \nabla u(x) \in \text{cspan}(J_g^T(x)), \forall x$$

If $u = f \circ g$

- Assume $u = f \circ g$ then $\nabla u(x) = J_g^T(x) J_f(g(x))$.Thus

$$\exists f \text{ s.t. } u = f \circ g \qquad \Longrightarrow \qquad \nabla u(x) \in \text{cspan}(J_g^T(x)), \forall x$$

- If $g^{-1}(z)$ is smoothly pathwise-connected for all $z \in \text{Im}(g)$,

$$\exists f \text{ s.t. } u = f \circ g \qquad \Longleftrightarrow \qquad \nabla u(x) \in \text{cspan}(J_g^T(x)), \forall x$$

If $u = f \circ g$

- Assume $u = f \circ g$ then $\nabla u(x) = J_g^T(x) J_f(g(x))$.Thus

$$\exists f \text{ s.t. } u = f \circ g \qquad \Longrightarrow \qquad \mathcal{L}(g) = 0$$

- If $g^{-1}(z)$ is smoothly pathwise-connected for all $z \in \mathrm{Im}(g)$,

$$\exists f \text{ s.t. } u = f \circ g \qquad \Longleftrightarrow \qquad \mathcal{L}(g) = 0$$

- Define

$$\mathcal{L}(g) := \mathbb{E}\left[\|\nabla u(X) - P_{\mathsf{cspan} J_g^T(X)} \nabla u(X)\|^2\right]$$

$$= \mathbb{E}\left[\nabla u^T \nabla u - \nabla u^T J_g^T (J_g J_g^T)^{-1} J_g \nabla u\right]$$

Poincaré Inequality

Assuming that a.s. $J_g(X)$ has rank m then we can write

$$\mathcal{L}(g) := \mathbb{E}\left[\|\nabla u_{|\mathcal{M}_Z}(X)\|^2\right],$$

where $\mathcal{M}_z := g^{-1}(z)$ is a Riemannian submanifold of $\mathbb{R}^d$ endowed with euclidean metric.

Definition (Poincaré Inequality)

Given $z \in Im(g)$, let $C(\mu_{X|Z=z})$ the smallest constant such that for any $h \in \mathcal{C}^1(\mathcal{M}_z, \mathbb{R})$,

$$\mathbb{E}\left[(h(X) - \mathbb{E}\left[h(X)\right])^2 | Z = z\right] \leq C(\mu_{X|Z=z}) \mathbb{E}\left[\|\nabla h(X)\|^2 | Z = z\right],$$

If $C(\mu_{X|Z=z}) < \infty$ we say that $\mu_{X|Z=z}$ satisfies a Poincaré Inequality.

Poincaré Inequality

Assuming that a.s. $J_g(X)$ has rank m then we can write

$$\mathcal{L}(g) := \mathbb{E}\left[\|\nabla u_{|\mathcal{M}_Z}(X)\|^2\right],$$

where $\mathcal{M}_z := g^{-1}(z)$ is a Riemannian submanifold of $\mathbb{R}^d$ endowed with euclidean metric.

Definition (Poincaré Inequality)

Given $z \in Im(g)$, let $C(\mu_{X|Z=z})$ the smallest constant such that for any $h \in \mathcal{C}^1(\mathcal{M}_z, \mathbb{R})$,

$$\mathbb{E}\left[(h(X) - \mathbb{E}\left[h(X)\right])^2 | Z = z\right] \leq C(\mu_{X|Z=z})\mathbb{E}\left[\|\nabla h(X)\|^2 | Z = z\right],$$

If $C(\mu_{X|Z=z}) < \infty$ we say that $\mu_{X|Z=z}$ satisfies a Poincaré Inequality.

Bounding the best error

(1) Assume $\text{rank}(J_g(X)) = m$ a.s., for all $g \in \mathcal{G}_m$.
(2) Assume $C(\mu_X, \mathcal{G}_m) < \infty$ where

$$C(\mu_X, \mathcal{G}_m) := \sup_{g \in \mathcal{G}_m} \sup_{z \in Im(g)} C(\mu_{X|Z=z}).$$

Proposition ([BMPZ22])

Under assumptions (1) and (2), it holds

$$\mathbb{E}\left[(u(X) - f_g \circ g(X))^2\right] \leq C(\mu_X, \mathcal{G}_m)\mathcal{L}(g)$$

We now focus on

$$\min_{g \in \mathcal{G}_m} \mathcal{L}(g)$$

Bounding the best error

(1) Assume $\text{rank}(J_g(X)) = m$ a.s., for all $g \in \mathcal{G}_m$.

(2) Assume $C(\mu_X, \mathcal{G}_m) < \infty$ where

$$C(\mu_X, \mathcal{G}_m) := \sup_{g \in \mathcal{G}_m} \sup_{z \in Im(g)} C(\mu_{X|Z=z}).$$

Proposition ([BMPZ22])

Under assumptions (1) and (2), it holds

$$\mathbb{E}\left[(u(X) - f_g \circ g(X))^2\right] \leq C(\mu_X, \mathcal{G}_m)\mathcal{L}(g)$$

We now focus on

$$\min_{g \in \mathcal{G}_m} \mathcal{L}(g)$$

Bounding the best error

(1) Assume $\text{rank}(J_g(X)) = m$ a.s., for all $g \in \mathcal{G}_m$.

(2) Assume $C(\mu_X, \mathcal{G}_m) < \infty$ where

$$C(\mu_X, \mathcal{G}_m) := \sup_{g \in \mathcal{G}_m} \sup_{z \in Im(g)} C(\mu_{X|Z=z}).$$

Proposition ([BMPZ22])

Under assumptions (1) and (2), it holds

$$\mathbb{E}\left[(u(X) - f_g \circ g(X))^2\right] \leq C(\mu_X, \mathcal{G}_m)\mathcal{L}(g)$$

We now focus on

$$\min_{g \in \mathcal{G}_m} \mathcal{L}(g)$$

Caveats:

- For general classes $\mathcal{G}_m$ bounding $C(\mu_X, \mathcal{G}_m)$ is an open problem.
- Worse : if $g^{-1}(z)$ is not connected then $C(\mu_{X|Z=z}) = \infty$.

Hopes:

- In practical cases the Poincaré bound is very pessimistic.
- Numerical experiments often showed a sharper bound.

The Poincaré Constant

Caveats:

- For general classes $\mathcal{G}_m$ bounding $C(\mu_X, \mathcal{G}_m)$ is an open problem.
- Worse : if $g^{-1}(z)$ is not connected then $C(\mu_{X|Z=z}) = \infty$.

Hopes:

- In practical cases the Poincaré bound is very pessimistic.
- Numerical experiments often showed a sharper bound.

The class $\mathcal{G}_m$

- [CDW14] Linear $g(x) = A^T x$ with $A \in \mathbb{R}^{d\times m}$.
  - $+$ Known bounds on $C(\mu_X, \mathcal{G}_m)$ for some classical μ_X.
  - $+$ Easy to minimize $\mathcal{L}$, i.e. to find the best A.
  - $-$ Restricted class.
- [VPZ23] Diffeomorphism-based $g(x) = (\psi_1(x), \cdots, \psi_m(x))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism .
  - $+$ Allow penalization for better control on $C(\mu_{X|Z=z})$.
  - $-$ Learning ψ can be difficult.
- [BMPZ22] Linear in features $g(x) = G^T \phi(x)$ with $\phi : \mathbb{R}^d \to \mathbb{R}^K$ and $G \in \mathbb{R}^{K\times m}$ with $K \geq d$.
  - $+$ Learning G is more reasonable.
  - - Cannot say much on $C(\mu_{X|Z=z})$.

The class $\mathcal{G}_m$

- [CDW14] Linear $g(x) = A^T x$ with $A \in \mathbb{R}^{d\times m}$.
  - + Known bounds on $C(\mu_X, \mathcal{G}_m)$ for some classical μ_X.
  - + Easy to minimize $\mathcal{L}$, i.e. to find the best A.
  - – Restricted class.
- [VPZ23] Diffeomorphism-based $g(x) = (\psi_1(x), \cdots, \psi_m(x))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism .
  - + Allow penalization for better control on $C(\mu_{X|Z=z})$.
  - – Learning ψ can be difficult.
- [BMPZ22] Linear in features $g(x) = G^T \phi(x)$ with $\phi : \mathbb{R}^d \to \mathbb{R}^K$ and $G \in \mathbb{R}^{K\times m}$ with $K \geq d$.
  - + Learning G is more reasonable.
  - - Cannot say much on $C(\mu_{X|Z=z})$.

The class $\mathcal{G}_m$

- [CDW14] Linear $g(x) = A^T x$ with $A \in \mathbb{R}^{d\times m}$.
  - $+$ Known bounds on $C(\mu_X, \mathcal{G}_m)$ for some classical μ_X.
  - $+$ Easy to minimize $\mathcal{L}$, i.e. to find the best A.
  - $-$ Restricted class.
- [VPZ23] Diffeomorphism-based $g(x) = (\psi_1(x), \cdots, \psi_m(x))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism .
  - $+$ Allow penalization for better control on $C(\mu_{X|Z=z})$.
  - $-$ Learning ψ can be difficult.
- [BMPZ22] Linear in features $g(x) = G^T \phi(x)$ with $\phi : \mathbb{R}^d \to \mathbb{R}^K$ and $G \in \mathbb{R}^{K\times m}$ with $K \geq d$.
  - $+$ Learning G is more reasonable.
  - - Cannot say much on $C(\mu_{X|Z=z})$.

The class $\mathcal{G}_m$

- [CDW14] Linear $g(x) = A^T x$ with $A \in \mathbb{R}^{d \times m}$.
  + Known bounds on $C(\mu_X, \mathcal{G}_m)$ for some classical μ_X.
  + Easy to minimize $\mathcal{L}$, i.e. to find the best A.
  – Restricted class.

- [VPZ23] Diffeomorphism-based $g(x) = (\psi_1(x), \cdots, \psi_m(x))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism .
  + Allow penalization for better control on $C(\mu_{X|Z=z})$.
  – Learning ψ can be difficult.

- [BMPZ22] Linear in features $g(x) = G^T \phi(x)$ with $\phi : \mathbb{R}^d \to \mathbb{R}^K$ and $G \in \mathbb{R}^{K \times m}$ with $K \geq d$.
  + Learning G is more reasonable.
  - Cannot say much on $C(\mu_{X|Z=z})$.

Minimizing $\mathcal{L}$

- Choose $\phi : \mathbb{R}^d \to \mathbb{R}^K$ with $K \geq d$, e.g. multivariate polynomials or splines.
- Tackle the following problem

$$\min_{\substack{G \in \mathbb{R}^{K \times m} \\ \text{constrains}(G)}} \quad \mathcal{L}(G^T \phi)$$

- [BMPZ22] Use a quasi-Newton method and go greedy on polynomial degree for ϕ.
- Questions :
  - Simpler way of building a good G ?
  - Other approach to build ϕ ?

Minimizing $\mathcal{L}$

- Choose $\phi : \mathbb{R}^d \to \mathbb{R}^K$ with $K \geq d$, e.g. multivariate polynomials or splines.
- Tackle the following problem

$$\min_{\substack{G \in \mathbb{R}^{K \times m} \\ \text{constrains}(G)}} \quad \mathcal{L}(G^T \phi)$$

- [BMPZ22] Use a quasi-Newton method and go greedy on polynomial degree for ϕ.
- Questions :
  - Simpler way of building a good G ?
  - Other approach to build ϕ ?

Minimizing $\mathcal{L}$

- Choose $\phi : \mathbb{R}^d \to \mathbb{R}^K$ with $K \geq d$, e.g. multivariate polynomials or splines.

- Tackle the following problem

$$\min_{\substack{G \in \mathbb{R}^{K \times m} \\ \text{constrains}(G)}} \mathcal{L}(G^T \phi)$$

- [BMPZ22] Use a quasi-Newton method and go greedy on polynomial degree for ϕ.

- Questions :
  - Simpler way of building a good G ?
  - Other approach to build ϕ ?

Minimizing $\mathcal{L}$

- Choose $\phi : \mathbb{R}^d \to \mathbb{R}^K$ with $K \geq d$, e.g. multivariate polynomials or splines.
- Tackle the following problem

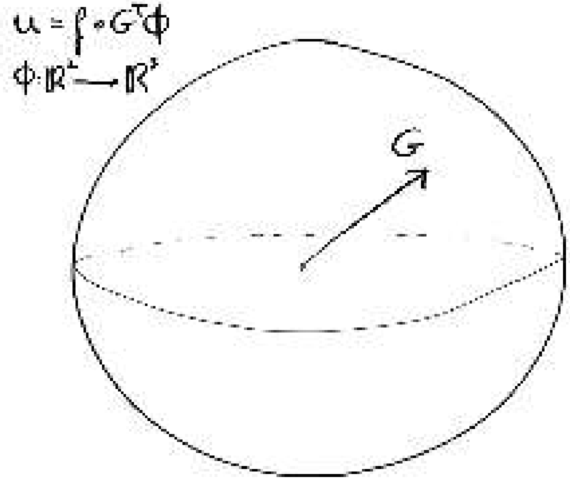
$$\min_{\substack{G \in \mathbb{R}^{K \times m} \\ \text{constrains}(G)}} \quad \mathcal{L}(G^T \phi)$$

- [BMPZ22] Use a quasi-Newton method and go greedy on polynomial degree for ϕ.
- Questions :
  - Simpler way of building a good G ?
  - Other approach to build ϕ ?

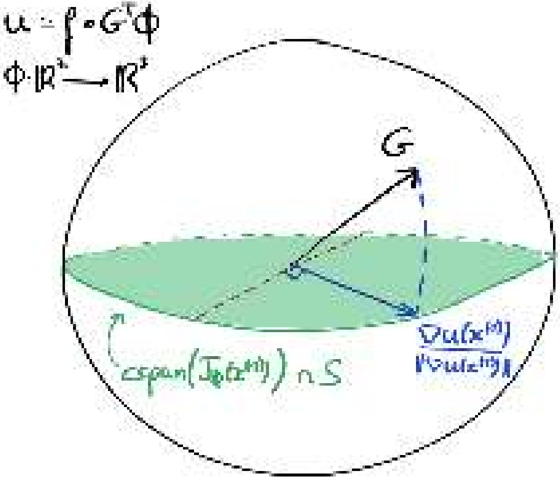
What if $u = f \circ G^T \phi$ and $m = 1$?

for some $\phi : \mathbb{R}^d \to \mathbb{R}^K$, some $G \in \mathbb{R}^{K \times m}$

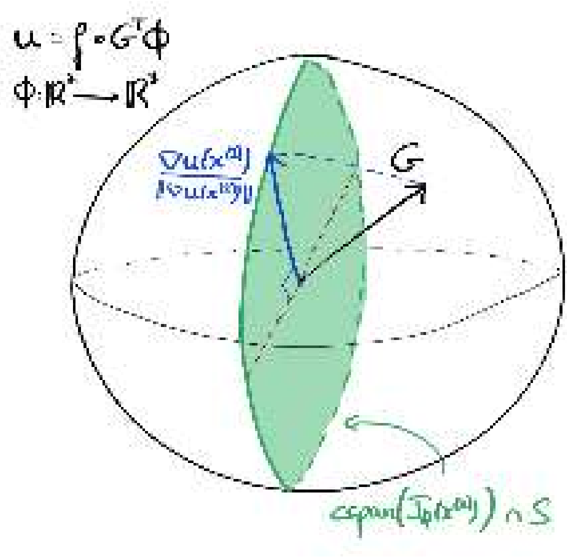
If $u = f \circ G^T \phi$ and $m = 1$



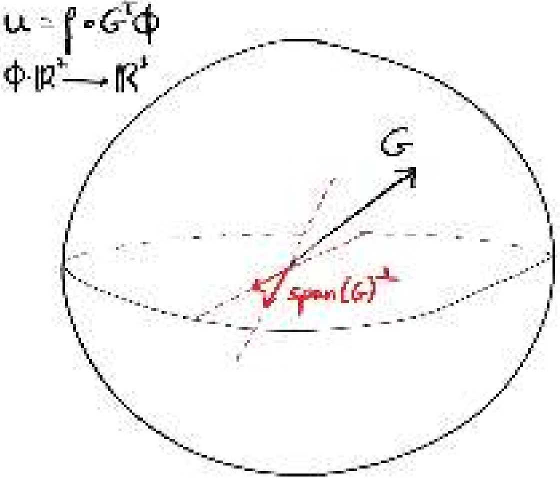
If $u = f \circ G^T \phi$ and $m = 1$



If $u = f \circ G^T \phi$ and $m = 1$



If $u = f \circ G^T \phi$ and $m = 1$



Case $m = 1$

- Define

$$\mathcal{J}(g) := \mathbb{E}\left[\|\nabla g(X) - P_{\mathsf{span}\nabla u(X)} \nabla g(X)\|^2\right].$$

which satisfies $\mathcal{J}(g) = 0 \iff \mathcal{L}(g) = 0$.

- Using $g = G^T \phi$ we can define some $H \in \mathbb{R}^{K \times K}$ s.t.

$$\mathcal{J}(g) = G^T H G.$$

$\rightarrow$ We can minimize $\mathcal{J}$ by solving an eigen-value problem !

$\rightarrow$ Problems :

- If $m > 1$ then $\mathcal{J}(g) = 0 \not\Leftrightarrow \mathcal{L}(g) = 0$
- If K gets too large ?

Case $m = 1$

- Define

$$\mathcal{J}(g) := \mathbb{E}\left[\|\nabla g(X) - P_{\mathsf{span}\nabla u(X)}\nabla g(X)\|^2\right].$$

which satisfies $\mathcal{J}(g) = 0 \iff \mathcal{L}(g) = 0$.

- Using $g = G^T\phi$ we can define some $H \in \mathbb{R}^{K\times K}$ s.t.

$$\mathcal{J}(g) = G^T H G.$$

$\rightarrow$ We can minimize $\mathcal{J}$ by solving an eigen-value problem !

$\rightarrow$ Problems :

- If $m > 1$ then $\mathcal{J}(g) = 0 \;\not\Leftrightarrow\; \mathcal{L}(g) = 0$
- If K gets too large ?

Case $m = 1$

- Define

$$\mathcal{J}(g) := \mathbb{E}\left[\|\nabla g(X) - P_{\mathsf{span}\nabla u(X)} \nabla g(X)\|^2\right].$$

which satisfies $\mathcal{J}(g) = 0 \iff \mathcal{L}(g) = 0$.

- Using $g = G^T \phi$ we can define some $H \in \mathbb{R}^{K \times K}$ s.t.

$$\mathcal{J}(g) = G^T H G.$$

$\rightarrow$ We can minimize $\mathcal{J}$ by solving an eigen-value problem !

$\rightarrow$ Problems :

- If $m > 1$ then $\mathcal{J}(g) = 0 \not\Longleftrightarrow \mathcal{L}(g) = 0$
- If K gets too large ?

Case $m = 1$

- Define

$$\mathcal{J}(g) := \mathbb{E}\left[\|\nabla g(X) - P_{\mathsf{span}\nabla u(X)}\nabla g(X)\|^2\right].$$

which satisfies $\mathcal{J}(g) = 0 \iff \mathcal{L}(g) = 0$.

- Using $g = G^T\phi$ we can define some $H \in \mathbb{R}^{K\times K}$ s.t.

$$\mathcal{J}(g) = G^T H G.$$

$\rightarrow$ We can minimize $\mathcal{J}$ by solving an eigen-value problem !

$\rightarrow$ Problems :

- If $m > 1$ then $\mathcal{J}(g) = 0 \not\Leftrightarrow \mathcal{L}(g) = 0$
- If K gets too large ?

Structured approach

Structured approach

- See $X = (S, T)$ consider g of the form

$$g(X) = G_3^T \phi_3\big(\, G_1^T \phi_1(S), G_2^T \phi_2(T)\,\big),$$

  for some chosen ϕ_i.

- First fit G_1 and G_2 independently, then fit G_3.

- Go further and consider a tree structure for g.

Structured approach

- See $X = (S, T)$ consider g of the form

$$g(X) = G_3^T \phi_3\big(G_1^T \phi_1(S), G_2^T \phi_2(T) \big),$$

  for some chosen ϕ_i.

- First fit G_1 and G_2 independently, then fit G_3.

- Go further and consider a tree structure for g.

Structured approach

- See $X = (S, T)$ consider g of the form

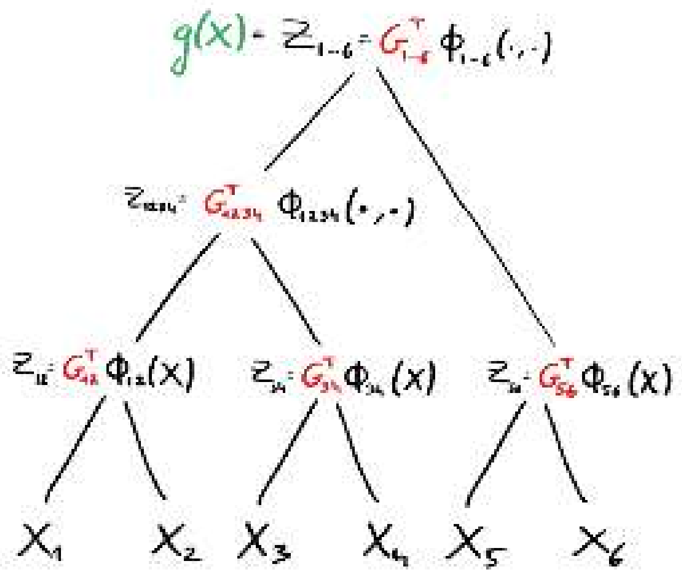
$$g(X) = G_3^T \phi_3\big(G_1^T \phi_1(S), G_2^T \phi_2(T) \big),$$

  for some chosen ϕ_i.

- First fit G_1 and G_2 independently, then fit G_3.

- Go further and consider a tree structure for g.

Structured approach



Structured approach

+ Better learning than Neural Nets while more expressive than Tensor Nets.

+ Several low-dim problems instead of one high-dim problem.

+ Under assumptions, we can use the approach from the case $m = 1$ to fit $G_1 \in \mathbb{R}^{K_1 \times m_1}$ and $G_2 \in \mathbb{R}^{K_2 \times m_2}$ with $m_1, m_2 > 1$.

- Good choice for the structure and for ϕ_i ?

Structured approach

+ Better learning than Neural Nets while more expressive than Tensor Nets.

+ Several low-dim problems instead of one high-dim problem.

+ Under assumptions, we can use the approach from the case $m = 1$ to fit $G_1 \in \mathbb{R}^{K_1 \times m_1}$ and $G_2 \in \mathbb{R}^{K_2 \times m_2}$ with $m_1, m_2 > 1$.

- Good choice for the structure and for ϕ_i ?

Structured approach

+ Better learning than Neural Nets while more expressive than Tensor Nets.

+ Several low-dim problems instead of one high-dim problem.

+ Under assumptions, we can use the approach from the case $m = 1$ to fit $G_1 \in \mathbb{R}^{K_1 \times m_1}$ and $G_2 \in \mathbb{R}^{K_2 \times m_2}$ with $m_1, m_2 > 1$.

- Good choice for the structure and for ϕ_i ?

Structured approach

+ Better learning than Neural Nets while more expressive than Tensor Nets.

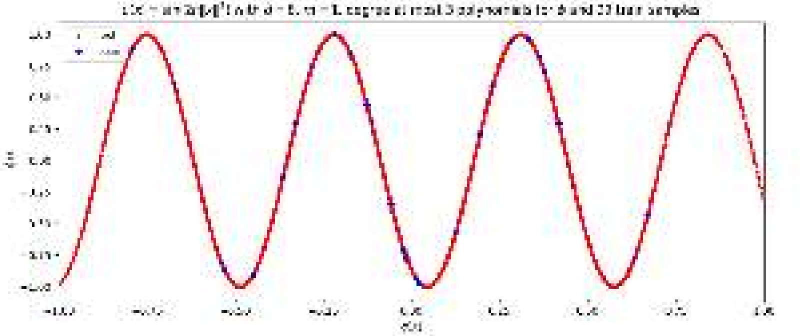
+ Several low-dim problems instead of one high-dim problem.

+ Under assumptions, we can use the approach from the case $m = 1$ to fit $G_1 \in \mathbb{R}^{K_1 \times m_1}$ and $G_2 \in \mathbb{R}^{K_2 \times m_2}$ with $m_1, m_2 > 1$.

- Good choice for the structure and for ϕ_i ?

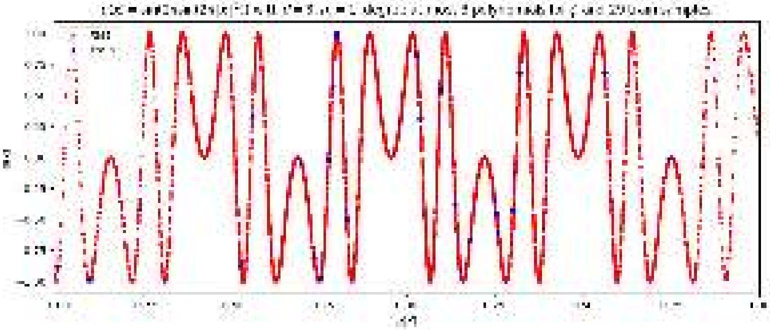
Numerical experiments

Structured approach



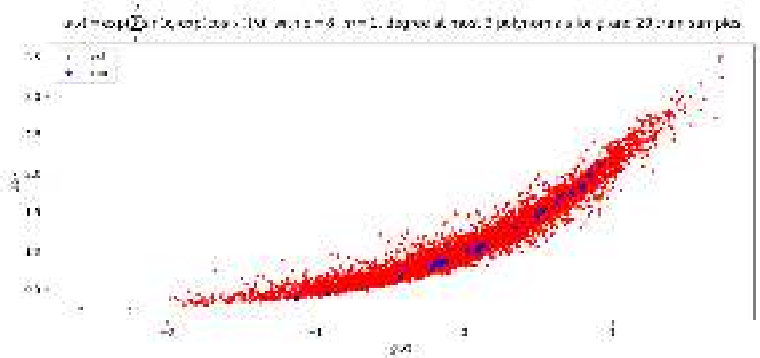
$$u(x) = \sin(2\pi\|x\|^2)$$

29 train samples



$$u(x) = \sin(2\pi \sin(2\pi \|x\|^2))$$

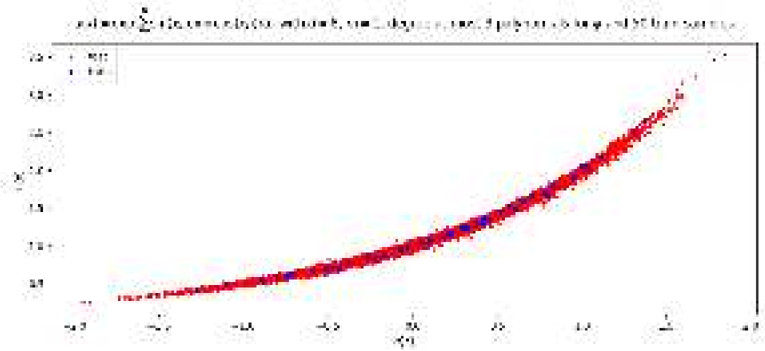
29 train samples



$$u(x) = \exp\big(\frac{1}{d}\sum_{i=1}^{d}\sin(x_i)\exp(\cos(x_i))\big)$$

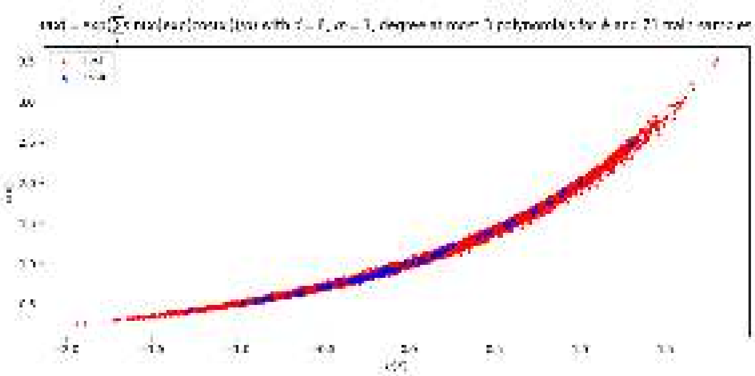
29 train samples

Structured approach



$$u(x) = \exp\Big(\frac{1}{d}\sum_{i=1}^{d}\sin(x_i)\exp(\cos(x_i))\Big)$$

50 train samples



$$u(x) = \exp\Big(\frac{1}{d}\sum_{i=1}^{d}\sin(x_i)\exp(\cos(x_i))\Big)$$

71 train samples

Conclusion and Perspectives

Conclusion and Perspectives

- Nonlinear dimension reduction method
- Structured way of fitting the feature map

$\rightarrow$ Investigate $\mathcal{L}(g)$ VS $\mathcal{J}(g)$

$\rightarrow$ Find conditions on G, ϕ and u so that a Poincaré-like Inequality holds.

$\rightarrow$ Investigate more challenging u.

Conclusion and Perspectives

- Nonlinear dimension reduction method
- Structured way of fitting the feature map

$\rightarrow$ Investigate $\mathcal{L}(g)$ VS $\mathcal{J}(g)$

$\rightarrow$ Find conditions on G, ϕ and u so that a Poincaré-like Inequality holds.

$\rightarrow$ Investigate more challenging u.

Thank you !

Appendix

References I

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, 11(4):1597–1639, December 2022.

Paul G. Constantine, Eric Dow, and Qiqi Wang.
Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces.
SIAM J. Sci. Comput., 36(4):A1500–A1524, January 2014.

Romain Verdière, Clémentine Prieur, and Olivier Zahm.
Diffeomorphism-based feature learning using Poincaré inequalities on augmented input space.
2023.

References II

Olivier Zahm, Paul G. Constantine, Clémentine Prieur, and Youssef M. Marzouk.
Gradient-Based Dimension Reduction of Multivariate Vector-Valued Functions.
SIAM J. Sci. Comput., 42(1):A534–A558, January 2020.