Structured Nonlinear Dimension Reduction Using Gradient Evaluations

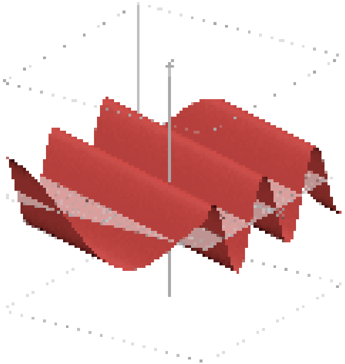
Alexandre PASCO

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

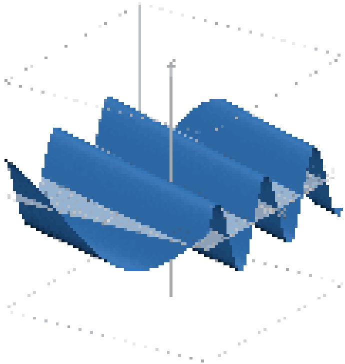
Introduction

Introduction

$u(x) = \sin(x_1 + 3x_2)$

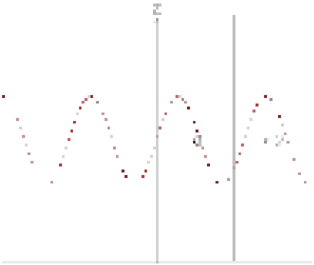


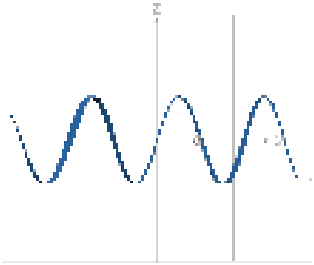
$v(x) = \sin(x_1 + 3x_2 + 0.05x_1x_2)$

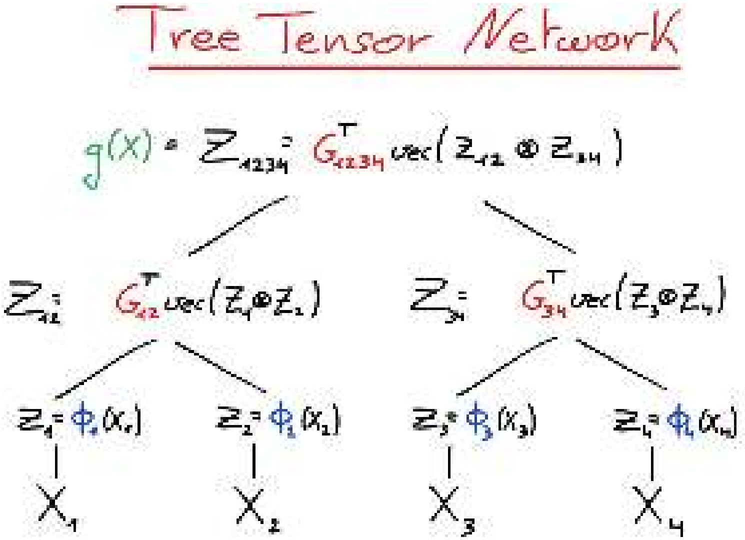


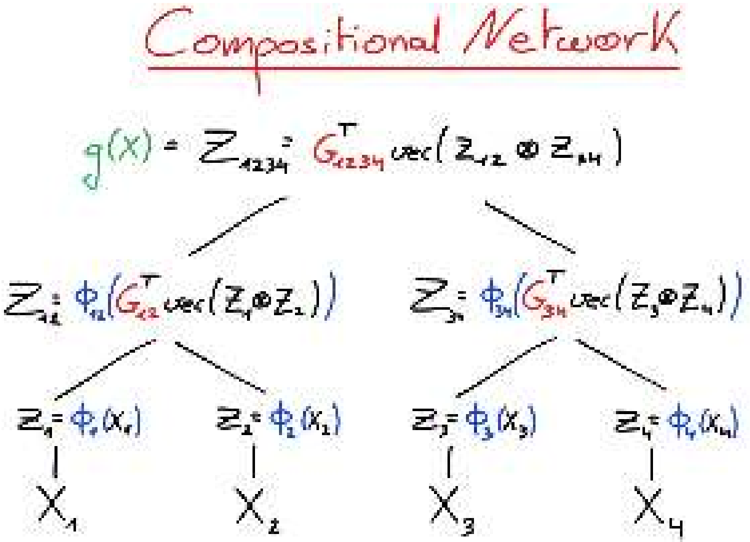
Introduction

$u(x) = \sin(x_1 + 3x_2)$ $\qquad$ $v(x) = \sin(x_1 + 3x_2 + 0.05x_1x_2)$






Tree Tensor Network
g(x) = Z1234 = G1234ᵀ vec(Z12 ⊗ Z34)
Z12 = G12ᵀ vec(Z1 ⊗ Z2)
Z34 = G34ᵀ vec(Z3 ⊗ Z4)
Z1 = Φ1(x1)
Z2 = Φ2(x2)
Z3 = Φ3(x3)
Z4 = Φ4(x4)
X1
X2
X3
X4


Compositional Network
g(x) = Z1234 = G1234^T vec(Z12 ⊗ Z34)
Z12 = Φ12(G12^T vec(Z1 ⊗ Z2))
Z34 = Φ34(G34^T vec(Z3 ⊗ Z4))
Z1 = Φ1(x1)
Z2 = Φ2(x2)
Z3 = Φ3(x3)
Z4 = Φ4(x4)
x1
x2
x3
x4

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

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

where X has probability density μ_X.

- Given : Few costly point evaluations

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

Introduction

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

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

where X has probability density μ_X.

- Given : Few costly point evaluations

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

Introduction

- Approximation of the form $\tilde{u} = f \circ g$.

- 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 $f : \mathbb{R}^m \to \mathbb{R}$.

Introduction

- Approximation of the form $\tilde{u} = f \circ g$.

- 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 $f : \mathbb{R}^m \to \mathbb{R}$.

Introduction

- Approximation of the form $\tilde{u} = f \circ g$.
- 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 $f : \mathbb{R}^m \to \mathbb{R}$.

- For a given g, the best profile map 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 **regression**,

$$f^* := \underset{f \in \mathcal{F}}{\operatorname{argmin}}\, \mathbb{E}\left[(u(X) - f(Z))^2\right]$$

- One can also consider gradient enhanced regression.

Introduction

- For a given g, the best profile map 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 **regression**,

$$f^* := \underset{f \in \mathcal{F}}{\operatorname{argmin}}\, \mathbb{E}\left[(u(X) - f(Z))^2\right]$$

- One can also consider gradient enhanced regression.

Introduction

- For a given g, the best profile map 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 **regression**,

$$f^* := \underset{f \in \mathcal{F}}{\operatorname{argmin}}\, \mathbb{E}\left[(u(X) - f(Z))^2\right]$$

- One can also consider gradient enhanced regression.

Measuring the quality of g

The class $\mathcal{G}_m$

- [Constantine et al., 2014] Linear $g(x) = A^T x$ with $A \in \mathbb{R}^{d\times m}$.
- [Bigoni et al., 2022] 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 $K \geq d$.
- [Verdière et al., 2023] Diffeomorphism-based $g(x) = (\psi_1(x), \cdots, \psi_m(x))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism.

The class $\mathcal{G}_m$

- [Constantine et al., 2014] Linear $g(x) = A^T x$ with $A \in \mathbb{R}^{d\times m}$.
- [Bigoni et al., 2022] 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 $K \geq d$.
- [Verdière et al., 2023] Diffeomorphism-based $g(x) = (\psi_1(x), \cdots, \psi_m(x))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism.

The class $\mathcal{G}_m$

- [Constantine et al., 2014] Linear $g(x) = A^T x$ with $A \in \mathbb{R}^{d \times m}$.
- [Bigoni et al., 2022] 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 $K \geq d$.
- [Verdière et al., 2023] Diffeomorphism-based $g(x) = (\psi_1(x), \cdots, \psi_m(x))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism.

Poincaré Inequality

- Assuming that a.s. $J_g(X)$ has rank m a.s., then $\mathcal{M}_z := g^{-1}(z)$ is a riemannian submanifold of $\mathbb{R}^d$.

- Let C_z the smallest constant such that for any $h \in \mathcal{C}^1(\mathcal{M}_z, \mathbb{R})$ with mean 0,

$$\mathbb{E}\left[h(X)^2 | Z = z\right] \leq C_z \mathbb{E}\left[\|\nabla h(X)\|^2 | Z = z\right].$$

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

- Take $h = (u - f_g \circ g)_{|\mathcal{M}_z}$, it is $\mathcal{C}^1$, has mean 0 and for $x \in \mathcal{M}_z$,

$$\|\nabla h(x)\|^2 = \|\nabla u(x)\|^2 - \|P_{\mathsf{span} J_g^T(x)} \nabla u(x)\|^2$$

Poincaré Inequality

- Assuming that a.s. $J_g(X)$ has rank m a.s., then $\mathcal{M}_z := g^{-1}(z)$ is a riemannian submanifold of $\mathbb{R}^d$.

- Let C_z the smallest constant such that for any $h \in \mathcal{C}^1(\mathcal{M}_z, \mathbb{R})$ with mean 0,

$$\mathbb{E}\left[h(X)^2 | Z = z\right] \leq C_z \mathbb{E}\left[\|\nabla h(X)\|^2 | Z = z\right].$$

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

- Take $h = (u - f_g \circ g)_{|\mathcal{M}_z}$, it is $\mathcal{C}^1$, has mean 0 and for $x \in \mathcal{M}_z$,

$$\|\nabla h(x)\|^2 = \|\nabla u(x)\|^2 - \|P_{\mathrm{span} J_g^T(x)} \nabla u(x)\|^2$$

Poincaré Inequality

- Assuming that a.s. $J_g(X)$ has rank m a.s., then $\mathcal{M}_z := g^{-1}(z)$ is a riemannian submanifold of $\mathbb{R}^d$.

- Let C_z the smallest constant such that for any $h \in \mathcal{C}^1(\mathcal{M}_z, \mathbb{R})$ with mean 0,

$$\mathbb{E}\left[h(X)^2 | Z = z\right] \leq C_z \mathbb{E}\left[\|\nabla h(X)\|^2 | Z = z\right].$$

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

- Take $h = (u - f_g \circ g)_{|\mathcal{M}_z}$, it is $\mathcal{C}^1$, has mean 0 and for $x \in \mathcal{M}_z$,

$$\|\nabla h(x)\|^2 = \|\nabla u(x)\|^2 - \|P_{\text{span} J_g^T(x)} \nabla u(x)\|^2$$

Poincaré-based upper-bound

For $g \in \mathcal{G}_m$ define

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

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

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

$$C(\mathcal{G}_m) := \sup_{g \in \mathcal{G}_m} \sup_{Z} C_Z.$$

Proposition ([Bigoni et al., 2022])

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

$$\min_{f:\mathbb{R}^m \to \mathbb{R}} \mathbb{E}\left[(u(X) - f \circ g(X))^2\right] \leq C(\mathcal{G}_m)\mathcal{J}(g)$$

Poincaré-based upper-bound

For $g \in \mathcal{G}_m$ define

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

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

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

$$C(\mathcal{G}_m) := \sup_{g \in \mathcal{G}_m} \sup_{Z} C_Z.$$

Proposition ([Bigoni et al., 2022])

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

$$\min_{f : \mathbb{R}^m \to \mathbb{R}} \mathbb{E}\left[(u(X) - f \circ g(X))^2\right] \leq C(\mathcal{G}_m) \mathcal{J}(g)$$

The Poincaré Constant

Caveats:

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

Hopes:

- More assumptions on u for better Poincaré-like bounds ?
- Numerical experiments have shown good performances.

The Poincaré Constant

Caveats:

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

Hopes:

- More assumptions on u for better Poincaré-like bounds ?
- Numerical experiments have shown good performances.

The class $\mathcal{G}_m$

- [Constantine et al., 2014] Linear $g(x) = A^T x$ with $A \in \mathbb{R}^{d \times m}$.
  - $+$ Known bounds on $C(\mathcal{G}_m)$ for some classical μ_X.
  - $+$ Easy to minimize $\mathcal{J}$, i.e. to find the best A.
  - $-$ Restricted class.
- [Bigoni et al., 2022] 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_z.
- [Verdière et al., 2023] 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_z.
  - $-$ Learning ψ can be difficult.

The class $\mathcal{G}_m$

- [Constantine et al., 2014] Linear $g(x) = A^T x$ with $A \in \mathbb{R}^{d\times m}$.
  - $+$ Known bounds on $C(\mathcal{G}_m)$ for some classical μ_X.
  - $+$ Easy to minimize $\mathcal{J}$, i.e. to find the best A.
  - $-$ Restricted class.
- [Bigoni et al., 2022] 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_z.
- [Verdière et al., 2023] 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_z.
  - $-$ Learning ψ can be difficult.

The class $\mathcal{G}_m$

- [Constantine et al., 2014] Linear $g(x) = A^T x$ with $A \in \mathbb{R}^{d \times m}$.
  - $+$ Known bounds on $C(\mathcal{G}_m)$ for some classical μ_X.
  - $+$ Easy to minimize $\mathcal{J}$, i.e. to find the best A.
  - $-$ Restricted class.

- [Bigoni et al., 2022] 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_z.

- [Verdière et al., 2023] 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_z.
  - $-$ Learning ψ can be difficult.

The class $\mathcal{G}_m$

- [Constantine et al., 2014] Linear $g(x) = A^T x$ with $A \in \mathbb{R}^{d \times m}$.
  - $+$ Known bounds on $C(\mathcal{G}_m)$ for some classical μ_X.
  - $+$ Easy to minimize $\mathcal{J}$, i.e. to find the best A.
  - $-$ Restricted class.

- [Bigoni et al., 2022] 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_z.

- [Verdière et al., 2023] 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_z.
  - $-$ Learning ψ can be difficult.

Learning G when $g = G^T \phi$

Learning G

- For a given $\phi : \mathbb{R}^d \to \mathbb{R}^K$ with $K \geq d$, minimizing the Poincaré bound means tackling

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

- [Bigoni et al., 2022] 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 ?

Learning G

- For a given $\phi : \mathbb{R}^d \to \mathbb{R}^K$ with $K \geq d$, minimizing the Poincaré bound means tackling

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

- [Bigoni et al., 2022] 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 ?

Learning G

- For a given $\phi : \mathbb{R}^d \to \mathbb{R}^K$ with $K \geq d$, minimizing the Poincaré bound means tackling

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

- [Bigoni et al., 2022] 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 ?

Learning G

- For a given $\phi : \mathbb{R}^d \to \mathbb{R}^K$ with $K \geq d$, minimizing the Poincaré bound means tackling

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

- [Bigoni et al., 2022] Use a quasi-Newton method and go greedy on polynomial degree for ϕ.

- Questions :
  - "Simpler" way of building a good G ? Eigen value problem !
  - Other approach to build ϕ ?

Learning G

- For a given $\phi : \mathbb{R}^d \to \mathbb{R}^K$ with $K \geq d$, minimizing the Poincaré bound means tackling

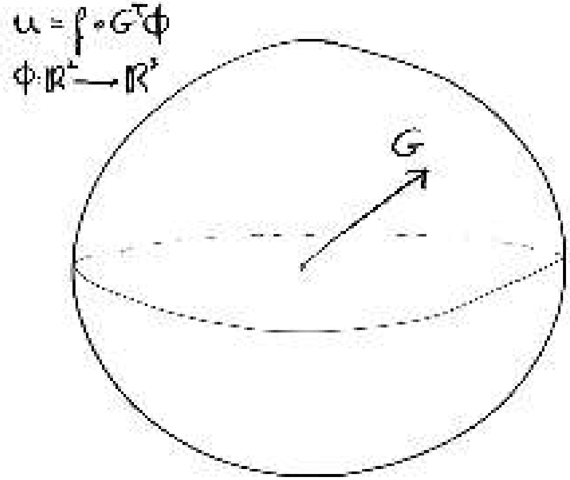
$$\min_{\substack{G \in \mathbb{R}^{K \times m} \\ \text{constrains}(G)}} \mathcal{J}(G^T \phi).$$

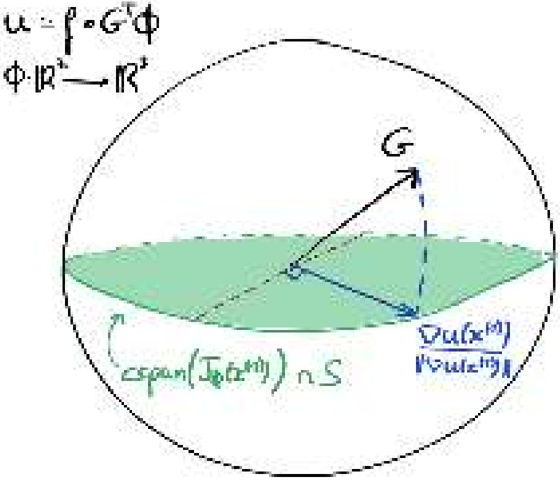
- [Bigoni et al., 2022] Use a quasi-Newton method and go greedy on polynomial degree for ϕ.
- Questions :
  - "Simpler" way of building a good G ? Eigen value problem !
  - Other approach to build ϕ ? Structured approach !

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

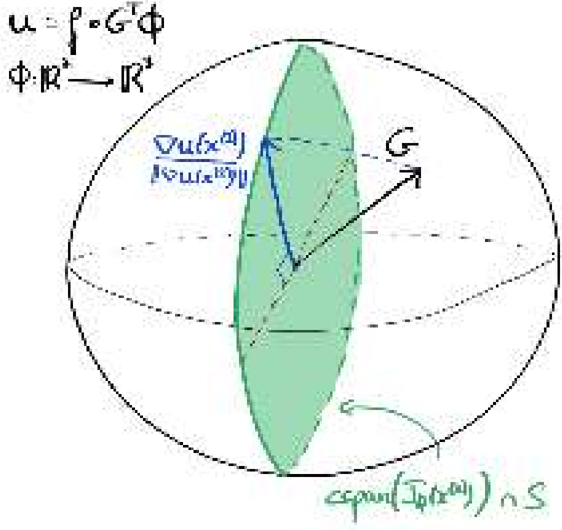
for some $\phi : \mathbb{R}^d \to \mathbb{R}^K$, some $G \in \mathbb{R}^{K \times m}$

Learning G: Geometric interpretation

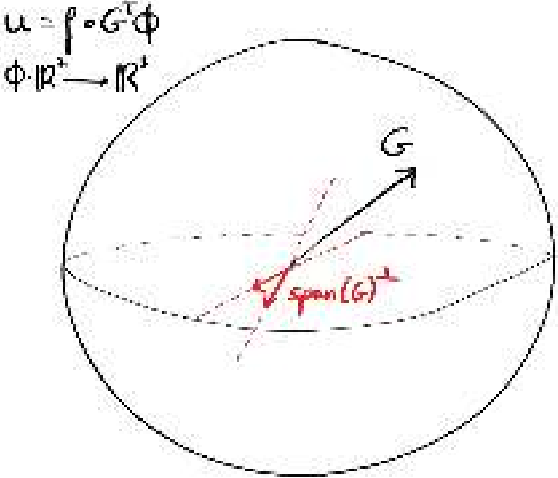



u = f ∘ G^T ϕ
G

Learning G: Geometric interpretation



Learning G: Geometric interpretation



Learning G: eigen value problem

- Assume $m = 1$ and define

$$\begin{aligned}\mathcal{L}(g) &:= \mathbb{E}\left[\|\nabla u\|^2 \|\nabla g - P_{\mathsf{span}\nabla u} \nabla g\|^2\right] \\ &= G^T H G\end{aligned}$$

where

$$H = \mathbb{E}\left[\|\nabla u\|^2 J_\phi J_\phi^T - J_\phi \nabla u \nabla u^T J_\phi^T\right] \in \mathbb{R}^{K \times K}$$

- If $m = 1$ then $\mathcal{L}(g) = 0 \iff \mathcal{J}(g) = 0$.

→ Minimizing $\mathcal{J}$ means solving an eigen-value problem !

→ Problems :
  - No realistic bound $\mathcal{L}(g) \leq \alpha \mathcal{J}(g)$ for all $g \in \mathcal{G}_m$.
  - If K gets too large ?
  - If $m > 1$?

Learning G: eigen value problem

- Assume $m = 1$ and define

$$\begin{aligned}\mathcal{L}(g) &:= \mathbb{E}\left[\|\nabla u\|^2 \|\nabla g - P_{\mathsf{span}\nabla u}\nabla g\|^2\right] \\ &= G^T H G\end{aligned}$$

where

$$H = \mathbb{E}\left[\|\nabla u\|^2 J_\phi J_\phi^T - J_\phi \nabla u \nabla u^T J_\phi^T\right] \in \mathbb{R}^{K\times K}$$

- If $m = 1$ then $\mathcal{L}(g) = 0 \iff \mathcal{J}(g) = 0$.

→ Minimizing $\mathcal{J}$ means solving an eigen-value problem !

→ Problems :

- No realistic bound $\mathcal{L}(g) \leq \alpha\mathcal{J}(g)$ for all $g \in \mathcal{G}_m$.
- If K gets too large ?
- If $m > 1$?

Learning G: eigen value problem

- Assume $m = 1$ and define

$$\begin{aligned}\mathcal{L}(g) &:= \mathbb{E}\left[\|\nabla u\|^2 \|\nabla g - P_{\mathsf{span}\nabla u} \nabla g\|^2\right] \\ &= G^T H G\end{aligned}$$

where

$$H = \mathbb{E}\left[\|\nabla u\|^2 J_\phi J_\phi^T - J_\phi \nabla u \nabla u^T J_\phi^T\right] \in \mathbb{R}^{K\times K}$$

- If $m = 1$ then $\mathcal{L}(g) = 0 \iff \mathcal{J}(g) = 0$.

$\rightarrow$ Minimizing $\mathcal{J}$ means solving an eigen-value problem !

$\rightarrow$ Problems :

- No realistic bound $\mathcal{L}(g) \leq \alpha \mathcal{J}(g)$ for all $g \in \mathcal{G}_m$.
- If K gets too large ?
- If $m > 1$?

Learning G: eigen value problem

- Assume $m = 1$ and define

$$\begin{aligned}\mathcal{L}(g) &:= \mathbb{E}\left[\|\nabla u\|^2 \|\nabla g - P_{\mathsf{span}\nabla u} \nabla g\|^2\right] \\ &= G^T H G\end{aligned}$$

where

$$H = \mathbb{E}\left[\|\nabla u\|^2 J_\phi J_\phi^T - J_\phi \nabla u \nabla u^T J_\phi^T\right] \in \mathbb{R}^{K \times K}$$

- If $m = 1$ then $\mathcal{L}(g) = 0 \iff \mathcal{J}(g) = 0$.

$\rightarrow$ Minimizing $\mathcal{J}$ means solving an eigen-value problem !

$\rightarrow$ Problems :

- No realistic bound $\mathcal{L}(g) \leq \alpha \mathcal{J}(g)$ for all $g \in \mathcal{G}_m$.
- If K gets too large ?
- If $m > 1$?

Learning G: eigen value problem

- Assume $m = 1$ and define

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

where

$$H = \mathbb{E}\left[\|\nabla u\|^2 J_\phi J_\phi^T - J_\phi \nabla u \nabla u^T J_\phi^T\right] \in \mathbb{R}^{K \times K}$$

- If $m = 1$ then $\mathcal{L}(g) = 0 \iff \mathcal{J}(g) = 0$.

→ Minimizing $\mathcal{J}$ means solving an eigen-value problem !

→ Problems :

- No realistic bound $\mathcal{L}(g) \leq \alpha \mathcal{J}(g)$ for all $g \in \mathcal{G}_m$.
- If K gets too large ? Add structure !
- If $m > 1$? Add structure !

Learning G: another eigen value problem

- Assume $m = 1$ and define

$$\mathcal{J}(g) := 1 - \mathbb{E}\left[\|\nabla u\|^2 \|P_{\mathsf{span}\nabla u} \nabla g\|^2\right] / \mathbb{E}\left[\|\nabla u\|^2 \|\nabla g\|^2\right]$$
$$= 1 - G^T H_1 G / G^T H_2 G$$

where

$$H_1 = \mathbb{E}\left[J_\phi \nabla u \nabla u^T J_\phi^T\right] \quad \text{and} \quad H_2 = \mathbb{E}\left[\|\nabla u\|^2 J_\phi J_\phi^T\right]$$

- If $m = 1$ then $\mathcal{L}(g) = 0 \iff \mathcal{J}(g) = 0$.

$\to$ Minimizing $\mathcal{J}$ means solving a generalized eigen-value problem !

$\to$ Problems :

- No realistic bound $\mathcal{J}(g) \leq \alpha \mathcal{J}(g)$ for all $g \in \mathcal{G}_m$.
- If K gets too large ? Add structure !
- If $m > 1$? Add structure !

Structured approach

Structured approach

- Consider $S = (X_1, \cdots, X_k)$, $T = (X_{k+1}, \cdots, X_d)$ and

$$g(X) = (g_1(S), g_2(T)) \in \mathbb{R}^{m_1+m_2}.$$

- Then g_1 and g_2 can be learnt separately by writing

$$\begin{aligned}\mathcal{J}(g) = \mathbb{E}\left[\|\nabla_S u(X)\|^2 - \|P_{\mathsf{span}\, J_{g_1}^T(S)} \nabla_S u(X)\|^2\right] \\ + \mathbb{E}\left[\|\nabla_T u(X)\|^2 - \|P_{\mathsf{span}\, J_{g_2}^T(T)} \nabla_T u(X)\|^2\right] \\ \mathcal{J}(g) = \mathcal{J}((g_1, \mathsf{id}_T)) + \mathcal{J}((\mathsf{id}_S, g_2))\end{aligned} \tag{1}$$

- Consider $\tilde{g}(X) = g_3(g_1(S), g_2(T))$ and learn g_3 minimizing $\mathcal{J}(\tilde{g})$.

Structured approach

- Consider $S = (X_1, \cdots, X_k)$, $T = (X_{k+1}, \cdots, X_d)$ and

$$g(X) = (g_1(S), g_2(T)) \in \mathbb{R}^{m_1+m_2}.$$

- Then g_1 and g_2 can be learnt separately by writing

$$\begin{aligned}\mathcal{J}(g) &= \mathbb{E}\left[\|\nabla_S u(X)\|^2 - \|P_{\mathsf{span} J_{g_1}^T(S)} \nabla_S u(X)\|^2\right] \\ &\quad + \mathbb{E}\left[\|\nabla_T u(X)\|^2 - \|P_{\mathsf{span} J_{g_2}^T(T)} \nabla_T u(X)\|^2\right] \\ \mathcal{J}(g) &= \mathcal{J}((g_1, \mathsf{id}_T)) + \mathcal{J}((\mathsf{id}_S, g_2))\end{aligned} \tag{1}$$

- Consider $\tilde{g}(X) = g_3(g_1(S), g_2(T))$ and learn g_3 minimizing $\mathcal{J}(\tilde{g})$.

Structured approach

- Consider $S = (X_1, \cdots, X_k)$, $T = (X_{k+1}, \cdots, X_d)$ and

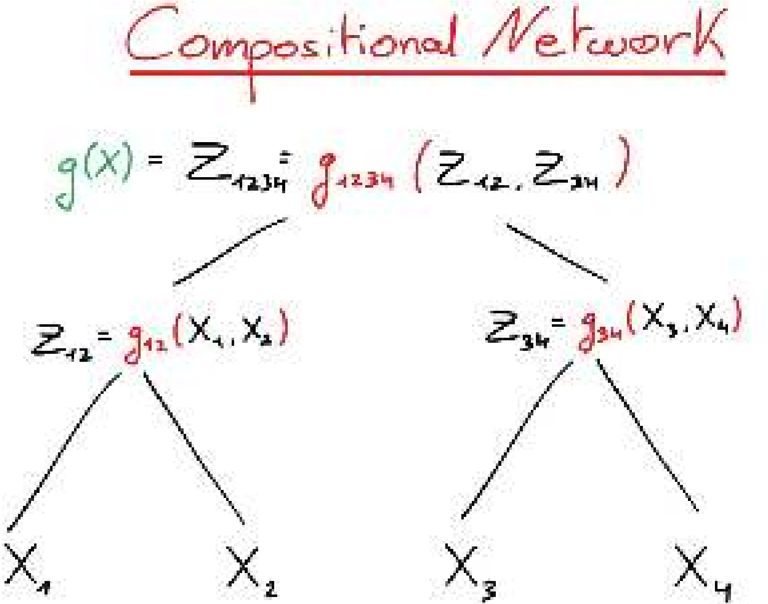
$$g(X) = (g_1(S), g_2(T)) \in \mathbb{R}^{m_1+m_2}.$$

- Then g_1 and g_2 can be learnt separately by writing

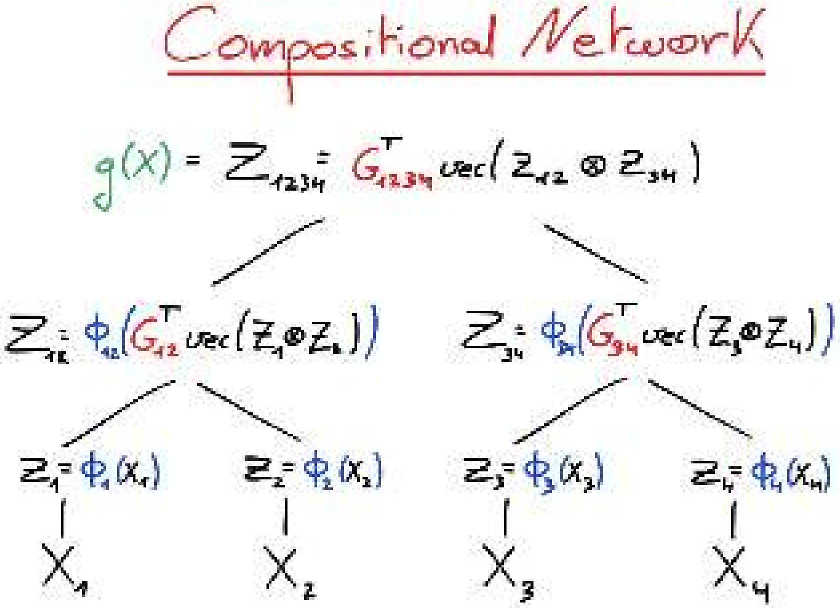
$$\begin{aligned}\mathcal{J}(g) &= \mathbb{E}\left[\|\nabla_S u(X)\|^2 - \|P_{\mathsf{span} J_{g_1}^T(S)} \nabla_S u(X)\|^2\right] \\ &\quad + \mathbb{E}\left[\|\nabla_T u(X)\|^2 - \|P_{\mathsf{span} J_{g_2}^T(T)} \nabla_T u(X)\|^2\right] \\ \mathcal{J}(g) &= \mathcal{J}((g_1, \mathsf{id}_T)) + \mathcal{J}((\mathsf{id}_S, g_2))\end{aligned} \tag{1}$$

- Consider $\tilde{g}(X) = g_3(g_1(S), g_2(T))$ and learn g_3 minimizing $\mathcal{J}(\tilde{g})$.

Structured approach



Structured approach



Structured approach

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

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

+ Under assumptions we can use the eig. val. formulation to learn $G_1 \in \mathbb{R}^{K_1 \times m_1}$ and $G_2 \in \mathbb{R}^{K_2 \times m_2}$ where $m_1, m_2 > 1$.

- Good choice for the structure and for ϕ_i ?

- There is still f to learn...

Structured approach

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

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

+ Under assumptions we can use the eig. val. formulation to learn $G_1 \in \mathbb{R}^{K_1 \times m_1}$ and $G_2 \in \mathbb{R}^{K_2 \times m_2}$ where $m_1, m_2 > 1$.

- Good choice for the structure and for ϕ_i ?

- There is still f to learn...

Conclusion and Perspectives

Conclusion and Perspectives

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

$\rightarrow$ Investigate $\min \mathcal{J}$ VS $\min \mathcal{J}$ and $\operatorname{argmin} \mathcal{J}$ VS $\operatorname{argmin} \mathcal{J}$

$\rightarrow$ Optimal sampling in this framework.

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

$\rightarrow$ Investigate challenging u, i.e. poorly approximated by TN and poorly learnt by NN.

Conclusion and Perspectives

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

→ Investigate $\min \mathcal{J}$ VS $\min \mathcal{J}$ and $\operatorname{argmin} \mathcal{J}$ VS $\operatorname{argmin} \mathcal{J}$

→ Optimal sampling in this framework.

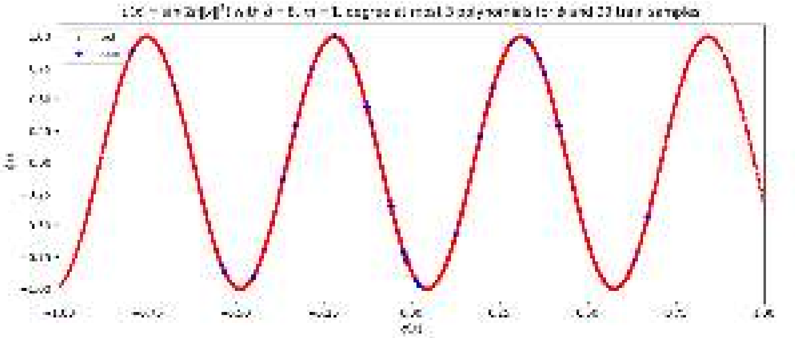
→ Find conditions on G, ϕ and u so that a Poincaré-like Inequality holds.

→ Investigate challenging u, i.e. poorly approximated by TN and poorly learnt by NN.

Thank you !

Numerical experiments

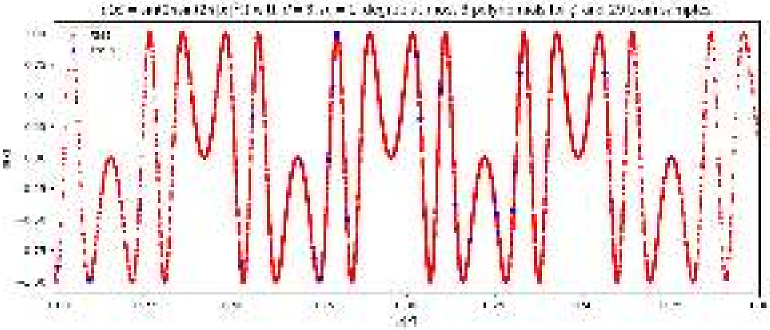
Numerical experiments



$$u(x) = \sin(2\pi\|x\|^2), \quad d = 8$$

29 train samples

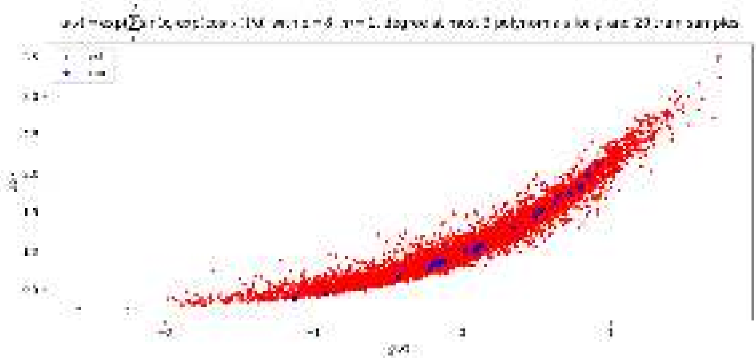
Numerical experiments



$$u(x) = \sin(2\pi \sin(2\pi \|x\|^2)), \quad d = 8$$

29 train samples

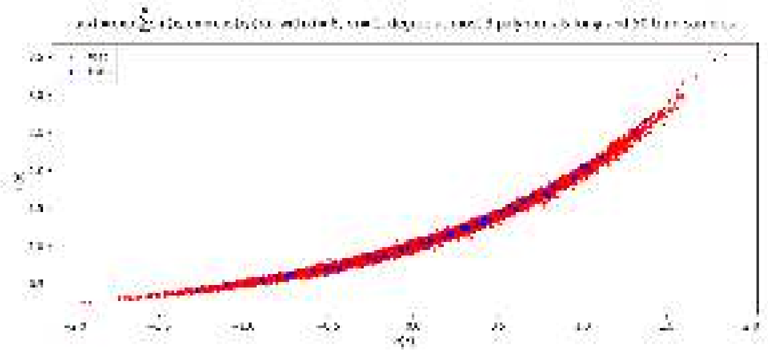
Numerical experiments



$$u(x) = \exp\left(\frac{1}{d}\sum_{i=1}^{d}\sin(x_i)\exp(\cos(x_i))\right), \quad d = 8$$

29 train samples

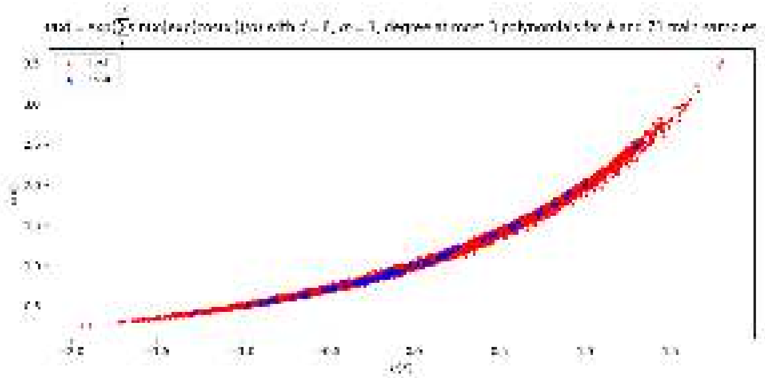
Numerical experiments



$$u(x) = \exp\big(\frac{1}{d}\sum_{i=1}^{d}\sin(x_i)\exp(\cos(x_i))\big), \quad d = 8$$

50 train samples

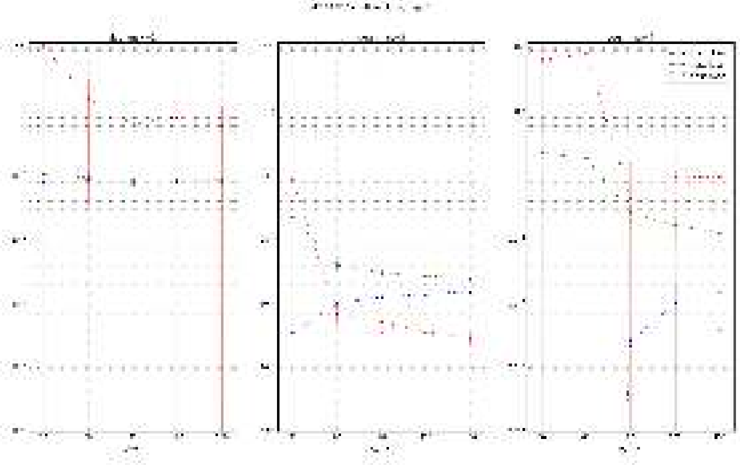
Numerical experiments



$$u(x) = \exp\big(\frac{1}{d}\sum_{i=1}^{d}\sin(x_i)\exp(\cos(x_i))\big), \quad d = 8$$

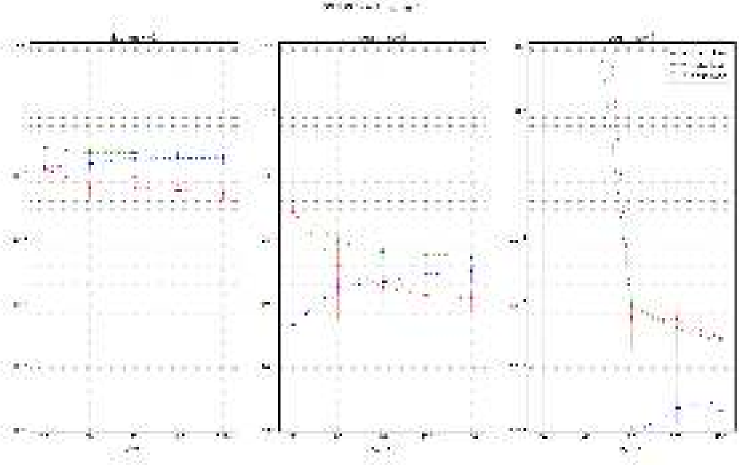
71 train samples

Numerical experiments



$$u(x) = \exp\Big(\frac{1}{d}\sum_{i=1}^{d}\sin(x_i)\exp(\cos(x_i))\Big), \quad d = 8$$

Numerical experiments



$$u(x) = \frac{2\pi X_2(X_6 - X_7)}{\ln(X_5/X_1)(1 + \frac{2X_4X_2}{\ln(X_5/X_1)X_1^2X_8} + \frac{X_2}{X_3})}, \quad d = 8$$

Appendix

References I

Bigoni, D., Marzouk, Y., Prieur, C., and Zahm, O. (2022).
Nonlinear dimension reduction for surrogate modeling using gradient information.
Information and Inference: A Journal of the IMA, 11(4):1597–1639.

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

Verdière, R., Prieur, C., and Zahm, O. (2023).
Diffeomorphism-based feature learning using Poincaré inequalities on augmented input space.
preprint.