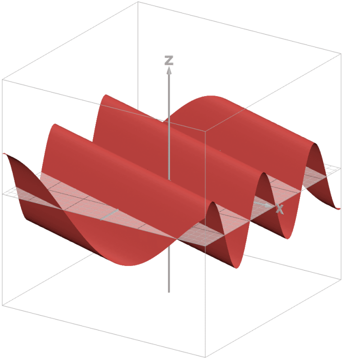
Surrogate to Poincaré inequalities on manifolds for dimension reduction in nonlinear feature spaces

Alexandre PASCO

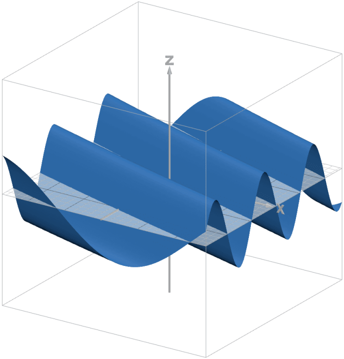
Centrale Nantes, Nantes Université, France
In collaboration with Anthony NOUY.

Introduction

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

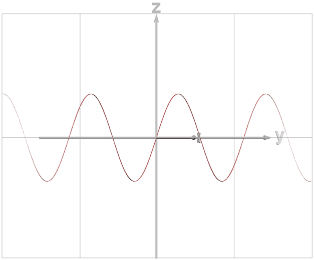


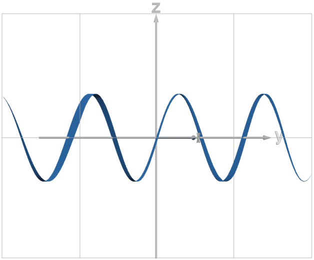
$v(x) = \sin(x_1+3x_2+0.05x_1x_2)$



Introduction

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





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

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

  where $\mathbf{X}$ has probability density $\mu_{\mathbf{X}}$.

- Given : Few costly point evaluations

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

Introduction

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

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

where $\mathbf{X}$ has probability density $\mu_{\mathbf{X}}$.

- Given : Few costly point evaluations

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

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

- 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

- For a given g, the best profile map is

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

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

- In practice : learn f^* via **regression**,

$$\inf_{f \in \mathcal{F}} \mathbb{E}\left[(u(\mathbf{X}) - f(\mathbf{Z}))^2\right]$$

- One can also consider gradient enhanced regression.

Introduction

- For a given g, the best profile map is

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

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

- In practice : learn f^* via **regression**,

$$\inf_{f\in\mathcal{F}} \mathbb{E}\left[(u(\mathbf{X}) - f(\mathbf{Z}))^2\right]$$

- One can also consider gradient enhanced regression.

- For a given g, the best profile map is

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

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

- In practice : learn f^* via **regression**,

$$\inf_{f \in \mathcal{F}} \mathbb{E}\left[(u(\mathbf{X}) - f(\mathbf{Z}))^2\right]$$

- One can also consider gradient enhanced regression.

The set $\mathcal{G}_m$

- [Constantine et al., 2014]
  Linear $g(\mathbf{x}) = G^T\mathbf{x}$ with $G \in \mathbb{R}^{d\times m}$.

- [Bigoni et al., 2022, Romor et al., 2022]
  Vector space $g(\mathbf{x}) = G^T\Phi(\mathbf{x})$ 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, Zhang et al., 2019]
  Diffeomorphism-based $g(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_m(\mathbf{x}))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism.

The set $\mathcal{G}_m$

- [Constantine et al., 2014]
  Linear $g(\mathbf{x}) = G^T\mathbf{x}$ with $G \in \mathbb{R}^{d\times m}$.

- [Bigoni et al., 2022, Romor et al., 2022]
  Vector space $g(\mathbf{x}) = G^T\Phi(\mathbf{x})$ 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, Zhang et al., 2019]
  Diffeomorphism-based $g(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_m(\mathbf{x}))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism.

The set $\mathcal{G}_m$

- [Constantine et al., 2014]
  Linear $g(\mathbf{x}) = G^T\mathbf{x}$ with $G \in \mathbb{R}^{d\times m}$.

- [Bigoni et al., 2022, Romor et al., 2022]
  Vector space $g(\mathbf{x}) = G^T\Phi(\mathbf{x})$ 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, Zhang et al., 2019]
  Diffeomorphism-based $g(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_m(\mathbf{x}))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism.

Poincaré Inequality

- Assuming $\nabla g(\mathbf{X})$ has rank m a.s., then $\mathcal{M}_{\mathbf{Z}} := g^{-1}(\mathbf{Z})$ is a smooth submanifold of $\mathbb{R}^d$.

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

$$\mathbb{E}\left[h(\mathbf{X})^2 | \mathbf{Z} = \mathbf{z}\right] \leq C_{\mathbf{Z}} \mathbb{E}\left[\|\nabla h(\mathbf{X})\|_2^2 | \mathbf{Z} = \mathbf{z}\right].$$

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

- Apply to $h = (u - f_g \circ g)_{|\mathcal{M}_{\mathbf{z}}}$ for $\mathbf{x} \in \mathcal{M}_{\mathbf{z}}$,

$$\|\nabla h(\mathbf{x})\|_2^2 = \|\Pi^{\perp}_{\nabla g(\mathbf{x})} \nabla u(\mathbf{x})\|_2^2 = \|\nabla u(\mathbf{x})\|_2^2 - \|\Pi_{\nabla g(\mathbf{x})} \nabla u(\mathbf{x})\|_2^2$$

Poincaré Inequality

- Assuming $\nabla g(\mathbf{X})$ has rank m a.s., then $\mathcal{M}_{\mathbf{Z}} := g^{-1}(\mathbf{Z})$ is a smooth submanifold of $\mathbb{R}^d$.

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

$$\mathbb{E}\left[h(\mathbf{X})^2 | \mathbf{Z} = \mathbf{z}\right] \leq C_{\mathbf{Z}} \mathbb{E}\left[\|\nabla h(\mathbf{X})\|_2^2 | \mathbf{Z} = \mathbf{z}\right].$$

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

- Apply to $h = (u - f_g \circ g)_{|\mathcal{M}_{\mathbf{z}}}$ for $\mathbf{x} \in \mathcal{M}_{\mathbf{z}}$,

$$\|\nabla h(\mathbf{x})\|_2^2 = \|\Pi^{\perp}_{\nabla g(\mathbf{x})} \nabla u(\mathbf{x})\|_2^2 = \|\nabla u(\mathbf{x})\|_2^2 - \|\Pi_{\nabla g(\mathbf{x})} \nabla u(\mathbf{x})\|_2^2$$

Poincaré Inequality

- Assuming $\nabla g(\mathbf{X})$ has rank m a.s., then $\mathcal{M}_{\mathbf{Z}} := g^{-1}(\mathbf{Z})$ is a smooth submanifold of $\mathbb{R}^d$.

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

$$\mathbb{E}\left[h(\mathbf{X})^2 | \mathbf{Z} = \mathbf{z}\right] \leq C_{\mathbf{Z}} \mathbb{E}\left[\|\nabla h(\mathbf{X})\|_2^2 | \mathbf{Z} = \mathbf{z}\right].$$

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

- Apply to $h = (u - f_g \circ g)_{|\mathcal{M}_{\mathbf{z}}}$ for $\mathbf{x} \in \mathcal{M}_{\mathbf{z}}$,

$$\|\nabla h(\mathbf{x})\|_2^2 = \|\Pi^{\perp}_{\nabla g(\mathbf{x})} \nabla u(\mathbf{x})\|_2^2 = \|\nabla u(\mathbf{x})\|_2^2 - \|\Pi_{\nabla g(\mathbf{x})} \nabla u(\mathbf{x})\|_2^2$$

Poincaré-based upper-bound

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

$$\mathcal{J}(g) := \mathbb{E}\left[\|\Pi^{\perp}_{\nabla g(\mathbf{X})} \nabla u(\mathbf{X})\|_2^2\right]$$

(1) Assume rank$(\nabla g(\mathbf{x})) = m$ for all $\mathbf{x} \in \mathcal{X}$ and all $g \in \mathcal{G}_m$.

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

$$C(\mathbf{X}|\mathcal{G}_m) := \sup_{g \in \mathcal{G}_m} \sup_{\mathbf{z} \in g(\mathcal{X})} C_{\mathbf{z}}.$$

Proposition ([Bigoni et al., 2022])

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

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

Poincaré-based upper-bound

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

$$\mathcal{J}(g) := \mathbb{E}\left[\|\Pi^{\perp}_{\nabla g(\mathbf{X})} \nabla u(\mathbf{X})\|_2^2\right]$$

(1) Assume $\text{rank}(\nabla g(\mathbf{x})) = m$ for all $\mathbf{x} \in \mathcal{X}$ and all $g \in \mathcal{G}_m$.

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

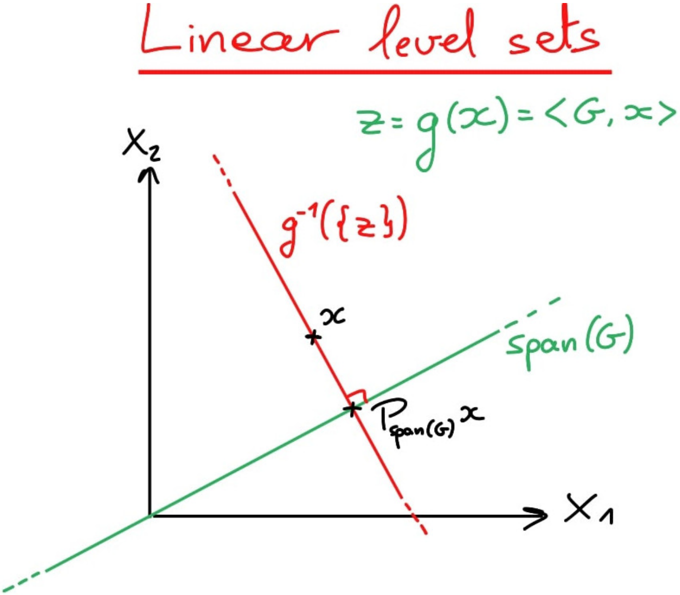
$$C(\mathbf{X}|\mathcal{G}_m) := \sup_{g \in \mathcal{G}_m} \sup_{\mathbf{z} \in g(\mathcal{X})} C_{\mathbf{z}}.$$

Proposition ([Bigoni et al., 2022])

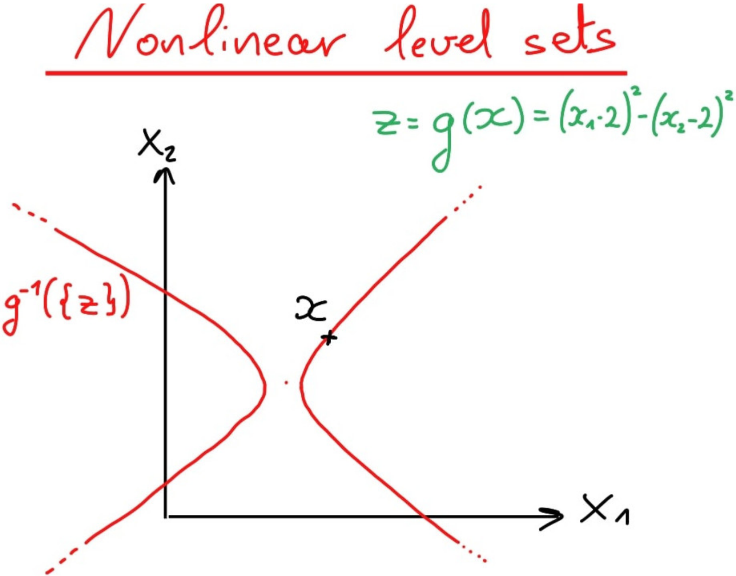
Under assumptions (1) and (2), it holds

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

Poincaré-based upper-bound



Poincaré-based upper-bound



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:

- Poincaré-like bounds are pessimistic.
- 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:

- Poincaré-like bounds are pessimistic.
- Numerical experiments have shown good performances.

The function $\mathcal{J}$

- Recall that

$$\mathcal{J}(g) = \mathbb{E}\left[\|\nabla u(\mathbf{X})\|_2^2\right] - \mathbb{E}\left[\|\Pi_{\nabla g(\mathbf{X})}\nabla u(\mathbf{X})\|_2^2\right]$$

- For $g(\mathbf{x}) = G^T\mathbf{x}$ with $G^TG = I_m$,

$$\Pi_{\nabla g(\mathbf{X})} = GG^T,$$

thus $G \mapsto \mathcal{J}(G^T)$ is **quadratic** and can be **explicitly minimized**.

- For $g(\mathbf{x}) = G^T\Phi(\mathbf{x})$ with fixed $\Phi : \mathbb{R}^d \to \mathbb{R}^K$,

$$\Pi_{\nabla g(\mathbf{X})} = \nabla\Phi(\mathbf{X})G(G^T\nabla\Phi(\mathbf{X})^T\nabla\Phi(\mathbf{X})G)^{-1}G^T\nabla\Phi(\mathbf{X})^T,$$

thus $\mathcal{J}$ is **not convex** anymore and **no explicit minimizer** is known.

The function $\mathcal{J}$

- Recall that

$$\mathcal{J}(g) = \mathbb{E}\left[\|\nabla u(\mathbf{X})\|_2^2\right] - \mathbb{E}\left[\|\Pi_{\nabla g(\mathbf{X})} \nabla u(\mathbf{X})\|_2^2\right]$$

- For $g(\mathbf{x}) = G^T\mathbf{x}$ with $G^TG = I_m$,

$$\Pi_{\nabla g(\mathbf{X})} = GG^T,$$

thus $G \mapsto \mathcal{J}(G^T)$ is **quadratic** and can be **explicitly minimized**.

- For $g(\mathbf{x}) = G^T\Phi(\mathbf{x})$ with fixed $\Phi : \mathbb{R}^d \to \mathbb{R}^K$,

$$\Pi_{\nabla g(\mathbf{X})} = \nabla\Phi(\mathbf{X})G(G^T\nabla\Phi(\mathbf{X})^T\nabla\Phi(\mathbf{X})G)^{-1}G^T\nabla\Phi(\mathbf{X})^T,$$

thus $\mathcal{J}$ is **not convex** anymore and **no explicit minimizer** is known.

The function $\mathcal{J}$

- Recall that

$$\mathcal{J}(g) = \mathbb{E}\left[\|\nabla u(\mathbf{X})\|_2^2\right] - \mathbb{E}\left[\|\Pi_{\nabla g(\mathbf{X})} \nabla u(\mathbf{X})\|_2^2\right]$$

- For $g(\mathbf{x}) = G^T\mathbf{x}$ with $G^TG = I_m$,

$$\Pi_{\nabla g(\mathbf{X})} = GG^T,$$

thus $G \mapsto \mathcal{J}(G^T)$ is **quadratic** and can be **explicitly minimized**.

- For $g(\mathbf{x}) = G^T\Phi(\mathbf{x})$ with fixed $\Phi : \mathbb{R}^d \to \mathbb{R}^K$,

$$\Pi_{\nabla g(\mathbf{X})} = \nabla\Phi(\mathbf{X})G(G^T\nabla\Phi(\mathbf{X})^T\nabla\Phi(\mathbf{X})G)^{-1}G^T\nabla\Phi(\mathbf{X})^T,$$

thus $\mathcal{J}$ is **not convex** anymore and **no explicit minimizer** is known.

The set $\mathcal{G}_m$

- Linear $g(\mathbf{x}) = G^T \mathbf{x}$ with $G \in \mathbb{R}^{d \times m}$.
  - $+$ Known bounds on $C(\mathbf{X}|\mathcal{G}_m)$ for some classical $\mu_{\mathbf{X}}$.
  - $+$ Easy to minimize $\mathcal{J}$, i.e. to find the best A.
  - $-$ Restricted class.
- Vector space $g(\mathbf{x}) = G^T \Phi(\mathbf{x})$ 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 ($\mathcal{J}$ non-convex but $\mathcal{G}_m$ convex).
  - \- Cannot say much on $C_{\mathbf{z}}$.
- Diffeomorphism-based $g(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_m(\mathbf{x}))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism .
  - $+$ Allow penalization for better control on $C_{\mathbf{z}}$.
  - $-$ Learning ψ can be difficult ($\mathcal{J}$ non-convex and $\mathcal{G}_m$ non-convex).

The set $\mathcal{G}_m$

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

- Vector space $g(\mathbf{x}) = G^T\Phi(\mathbf{x})$ 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 ($\mathcal{J}$ non-convex but $\mathcal{G}_m$ convex).
  - \- Cannot say much on $C_{\mathbf{z}}$.

- Diffeomorphism-based $g(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_m(\mathbf{x}))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism .
  - $+$ Allow penalization for better control on $C_{\mathbf{z}}$.
  - $-$ Learning ψ can be difficult ($\mathcal{J}$ non-convex and $\mathcal{G}_m$ non-convex).

The set $\mathcal{G}_m$

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

- Vector space $g(\mathbf{x}) = G^T\Phi(\mathbf{x})$ 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 ($\mathcal{J}$ non-convex but $\mathcal{G}_m$ convex).
  - \- Cannot say much on $C_{\mathbf{z}}$.

- Diffeomorphism-based $g(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_m(\mathbf{x}))^T$ where $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is a diffeomorphism .
  - $+$ Allow penalization for better control on $C_{\mathbf{z}}$.
  - $-$ Learning ψ can be difficult ($\mathcal{J}$ non-convex and $\mathcal{G}_m$ non-convex).

Surrogate one feature: definition

- The loss function $\mathcal{J}$ now writes

$$\mathcal{J}(g) = \mathbb{E}\left[\|\nabla u(\mathbf{X})\|_2^2 - \frac{(\nabla g(\mathbf{X})^T \nabla u(\mathbf{X}))^2}{\|\nabla g(\mathbf{X})\|_2^2}\right]$$
$$= \mathbb{E}\left[\frac{1}{\|\nabla g(\mathbf{X})\|_2^2} \underbrace{\|\nabla u(\mathbf{X})\|_2^2 \|\Pi^{\perp}_{\nabla u(\mathbf{X})} \nabla g(\mathbf{X})\|_2^2}_{\text{Quadratic wrt } g}\right]$$

- Controlling $\|\nabla g(\mathbf{X})\|_2^2$ allows to control $\mathcal{J}(g)$ by a **quadratic** surrogate,

$$\mathcal{L}_1(g) := \mathbb{E}\left[\|\nabla u(\mathbf{X})\|_2^2 \|\Pi^{\perp}_{\nabla u(\mathbf{X})} \nabla g(\mathbf{X})\|_2^2\right].$$

- Problem : control should be **uniform** on $\mathcal{G}_1$.

Surrogate one feature: definition

- The loss function $\mathcal{J}$ now writes

$$\mathcal{J}(g) = \mathbb{E}\left[\|\nabla u(\mathbf{X})\|_2^2 - \frac{(\nabla g(\mathbf{X})^T \nabla u(\mathbf{X}))^2}{\|\nabla g(\mathbf{X})\|_2^2}\right]$$
$$= \mathbb{E}\left[\frac{1}{\|\nabla g(\mathbf{X})\|_2^2}\underbrace{\|\nabla u(\mathbf{X})\|_2^2 \|\Pi^{\perp}_{\nabla u(\mathbf{X})} \nabla g(\mathbf{X})\|_2^2}_{\text{Quadratic wrt } g}\right]$$

- Controlling $\|\nabla g(\mathbf{X})\|_2^2$ allows to control $\mathcal{J}(g)$ by a **quadratic** surrogate,

$$\mathcal{L}_1(g) := \mathbb{E}\left[\|\nabla u(\mathbf{X})\|_2^2 \|\Pi^{\perp}_{\nabla u(\mathbf{X})} \nabla g(\mathbf{X})\|_2^2\right].$$

- Problem : control should be **uniform** on $\mathcal{G}_1$.

Surrogate one feature: definition

- The loss function $\mathcal{J}$ now writes

$$\mathcal{J}(g) = \mathbb{E}\left[\|\nabla u(\mathbf{X})\|_2^2 - \frac{(\nabla g(\mathbf{X})^T \nabla u(\mathbf{X}))^2}{\|\nabla g(\mathbf{X})\|_2^2}\right]$$
$$= \mathbb{E}\left[\frac{1}{\|\nabla g(\mathbf{X})\|_2^2} \underbrace{\|\nabla u(\mathbf{X})\|_2^2 \|\Pi^{\perp}_{\nabla u(\mathbf{X})} \nabla g(\mathbf{X})\|_2^2}_{\text{Quadratic wrt } g}\right]$$

- Controlling $\|\nabla g(\mathbf{X})\|_2^2$ allows to control $\mathcal{J}(g)$ by a **quadratic** surrogate,

$$\mathcal{L}_1(g) := \mathbb{E}\left[\|\nabla u(\mathbf{X})\|_2^2 \|\Pi^{\perp}_{\nabla u(\mathbf{X})} \nabla g(\mathbf{X})\|_2^2\right].$$

- Problem : control should be **uniform** on $\mathcal{G}_1$.

Surrogate one feature: Uniform control

- Uniformly bi-Lipschitz (ideal setting), i.e

$$\forall g \in \mathcal{G}_1, \quad 0 < c \leq \|\nabla g(\mathbf{X})\|_2^2 \leq C < +\infty$$

+ Possible when : linear or diffeomorphism-based.

– Not possible when : vector space of dimension $K > d$.

- Deviation inequalities, i.e rate of convergence of

$$\forall g \in \mathcal{G}_1, \quad \underbrace{\mathbb{P}\left[\|\nabla g(\mathbf{X})\|_2^2 \leq \beta^{-1}\right]}_{\text{Small deviations}}, \underbrace{\mathbb{P}\left[\|\nabla g(\mathbf{X})\|_2^2 \geq \beta\right]}_{\text{Large deviations}} \xrightarrow[\beta \to +\infty]{} 0$$

+ Possible when : $\mathbf{x} \mapsto \|\nabla g(\mathbf{x})\|_2^2$ satisfies a type of Remez inequality.

Surrogate one feature: Uniform control

- Uniformly bi-Lipschitz (ideal setting), i.e

$$\forall g \in \mathcal{G}_1, \quad 0 < c \leq \|\nabla g(\mathbf{X})\|_2^2 \leq C < +\infty$$

\+ Possible when : linear or diffeomorphism-based.

− Not possible when : vector space of dimension $K > d$.

- Deviation inequalities, i.e rate of convergence of

$$\forall g \in \mathcal{G}_1, \quad \underbrace{\mathbb{P}\left[\|\nabla g(\mathbf{X})\|_2^2 \leq \beta^{-1}\right]}_{\text{Small deviations}}, \underbrace{\mathbb{P}\left[\|\nabla g(\mathbf{X})\|_2^2 \geq \beta\right]}_{\text{Large deviations}} \xrightarrow[\beta \to +\infty]{} 0$$

\+ Possible when : $\mathbf{x} \mapsto \|\nabla g(\mathbf{x})\|_2^2$ satisfies a type of Remez inequality.

Surrogate one feature: Deviation inequalities

Consider $h : \mathbb{R}^d \to \mathbb{R}$. Assume that

(1) $\mathbf{X}$ has s-concave, $s \in (0, 1/d]$, proba law (implies compactly supported on a convex set).

(2) For any bounded line $J \subset \mathbb{R}^d$ and any measurable $I \subset J$,

$$\sup_{\mathbf{x} \in J} |h(\mathbf{x})| \leq \Big(\frac{A_h |J|}{|I|}\Big)^{k_h} \sup_{\mathbf{x} \in I} |h(\mathbf{x})|,$$

e.g. $A_h = 4$ and $k_h = k$ for polynomial with total degree $\leq k$.

Proposition (Direct consequence of [Fradelizi, 2009])

Under (1) and (2) and for some η_h, it holds for all $\beta > 0$,

$$\mathbb{P}\left[|h(\mathbf{X})| \leq \beta^{-1}\right] \lesssim \beta^{-1/k_h}, \quad \mathbb{P}\left[|h(\mathbf{X})| \geq \beta^{-1}\right] \lesssim 1_{\beta \leq \eta_h}.$$

For $s \in [-\infty, 0]$, still holds but slower decay for large deviations.

Surrogate one feature: Deviation inequalities

Consider $h : \mathbb{R}^d \to \mathbb{R}$. Assume that

(1) $\mathbf{X}$ has s-concave, $s \in (0, 1/d]$, proba law (implies compactly supported on a convex set).

(2) For any bounded line $J \subset \mathbb{R}^d$ and any measurable $I \subset J$,

$$\sup_{\mathbf{x}\in J} |h(\mathbf{x})| \leq \Big(\frac{A_h|J|}{|I|}\Big)^{k_h} \sup_{\mathbf{x}\in I} |h(\mathbf{x})|,$$

e.g. $A_h = 4$ and $k_h = k$ for polynomial with total degree $\leq k$.

Proposition (Direct consequence of [Fradelizi, 2009])

Under (1) and (2) and for some η_h, it holds for all $\beta > 0$,

$$\mathbb{P}\left[|h(\mathbf{X})| \leq \beta^{-1}\right] \lesssim \beta^{-1/k_h}, \quad \mathbb{P}\left[|h(\mathbf{X})| \geq \beta^{-1}\right] \lesssim 1_{\beta \leq \eta_h}.$$

For $s \in [-\infty, 0]$, still holds but slower decay for large deviations.

Surrogate one feature: Deviation inequalities

Consider $h : \mathbb{R}^d \to \mathbb{R}$. Assume that

(1) $\mathbf{X}$ has s-concave, $s \in (0, 1/d]$, proba law (implies compactly supported on a convex set).

(2) For any bounded line $J \subset \mathbb{R}^d$ and any measurable $I \subset J$,

$$\sup_{\mathbf{x}\in J} |h(\mathbf{x})| \leq \Big(\frac{A_h|J|}{|I|}\Big)^{k_h} \sup_{\mathbf{x}\in I} |h(\mathbf{x})|,$$

e.g. $A_h = 4$ and $k_h = k$ for polynomial with total degree $\leq k$.

Proposition (Direct consequence of [Fradelizi, 2009])

Under (1) and (2) and for some η_h, it holds for all $\beta > 0$,

$$\mathbb{P}\left[|h(\mathbf{X})| \leq \beta^{-1}\right] \lesssim \beta^{-1/k_h}, \quad \mathbb{P}\left[|h(\mathbf{X})| \geq \beta^{-1}\right] \lesssim 1_{\beta \leq \eta_h}.$$

For $s \in [-\infty, 0]$, still holds but slower decay for large deviations.

Surrogate one feature: Deviation inequalities

Consider $h : \mathbb{R}^d \to \mathbb{R}$. Assume that

(1) $\mathbf{X}$ has s-concave, $s \in (0, 1/d]$, proba law (implies compactly supported on a convex set).

(2) For any bounded line $J \subset \mathbb{R}^d$ and any measurable $I \subset J$,

$$\sup_{\mathbf{x} \in J} |h(\mathbf{x})| \leq \Big(\frac{A_h |J|}{|I|}\Big)^{k_h} \sup_{\mathbf{x} \in I} |h(\mathbf{x})|,$$

e.g. $A_h = 4$ and $k_h = k$ for polynomial with total degree $\leq k$.

Proposition (Direct consequence of [Fradelizi, 2009])

Under (1) and (2) and for some η_h, it holds for all $\beta > 0$,

$$\mathbb{P}\left[|h(\mathbf{X})| \leq \beta^{-1}\right] \lesssim \beta^{-1/k_h}, \quad \mathbb{P}\left[|h(\mathbf{X})| \geq \beta^{-1}\right] \lesssim 1_{\beta \leq \eta_h}.$$

For $s \in [-\infty, 0]$, still holds but slower decay for large deviations.

Surrogate one feature: Suboptimality result

- If $\mathcal{G}_1$ contains only non-constant polynomials of total degree at most $\ell+1$, then the previous deviations inequalities hold uniformly on $\mathcal{G}_1$ with $k = 2\ell$ and $A = 4$.

Proposition

Under (1), with $\mathcal{G}_1$ as above, it holds

$$\forall g \in \mathcal{G}_1, \quad \gamma_1 \mathcal{L}_1(g) \le \mathcal{J}(g) \le \gamma_2 \mathcal{L}_1(g)^{\frac{1}{1+2\ell}},$$

for some $0 < \gamma_1, \gamma_2 < +\infty$. In particular, for some $0 < \gamma_3 < +\infty$,

$$\mathcal{J}(g^*) \le \gamma_3 \inf_{\mathcal{G}_1} \mathcal{J}^{\frac{1}{1+2\ell}}$$

Similar results hold for $s \in [-\infty, 0]$.

Surrogate one feature: Suboptimality result

- If $\mathcal{G}_1$ contains only non-constant polynomials of total degree at most $\ell + 1$, then the previous deviations inequalities hold uniformly on $\mathcal{G}_1$ with $k = 2\ell$ and $A = 4$.

Proposition

Under (1)*, with* $\mathcal{G}_1$ *as above, it holds*

$$\forall g \in \mathcal{G}_1, \quad \gamma_1 \mathcal{L}_1(g) \leq \mathcal{J}(g) \leq \gamma_2 \mathcal{L}_1(g)^{\frac{1}{1+2\ell}},$$

for some $0 < \gamma_1, \gamma_2 < +\infty$*. In particular, for some* $0 < \gamma_3 < +\infty$*,*

$$\mathcal{J}(g^*) \leq \gamma_3 \inf_{\mathcal{G}_1} \mathcal{J}^{\frac{1}{1+2\ell}}$$

Similar results hold for $s \in [-\infty, 0]$.

Surrogate one feature: Vector space

- For assumptions to hold, we consider a sym pd $R \in \mathbb{R}^{K \times K}$,

$$\mathcal{G}_1 := \left\{ g : \mathbf{x} \mapsto G^T \Phi(\mathbf{x}) : G \in \mathbb{R}^K, G^T R G = 1 \right\},$$

- [Bigoni et al., 2022] Invariance property of $\mathcal{J}$ implies

$$\inf_{g \in \mathsf{span}\{\Phi_1, \cdots, \Phi_K\}} \mathcal{J}(g) = \inf_{g \in \mathcal{G}_1} \mathcal{J}(g)$$

Proposition

$$\min_{g \in \mathcal{G}_1} \mathcal{L}_1(g) = \min_{\substack{G \in \mathbb{R}^K \\ G^T R G = 1}} G^T H G,$$

where $H := H^{(1)} - H^{(2)} \in \mathbb{R}^{K \times K}$ *is sym psd and*

$$H^{(1)} := \mathbb{E}\left[\|\nabla u(\mathbf{X})\|_2^2 \nabla \Phi(\mathbf{X})^T \nabla \Phi(\mathbf{X}) \right] \in \mathbb{R}^{K \times K},$$
$$H^{(2)} := \mathbb{E}\left[\nabla \Phi(\mathbf{X})^T \nabla u(\mathbf{X}) \nabla u(\mathbf{X})^T \nabla \Phi(\mathbf{X}) \right] \in \mathbb{R}^{K \times K}.$$

Surrogate one feature: Vector space

- For assumptions to hold, we consider a sym pd $R \in \mathbb{R}^{K \times K}$,

$$\mathcal{G}_1 := \Big\{ g : \mathbf{x} \mapsto G^T \Phi(\mathbf{x}) : G \in \mathbb{R}^K, G^T R G = 1 \Big\},$$

- [Bigoni et al., 2022] Invariance property of $\mathcal{J}$ implies

$$\inf_{g \in \mathsf{span}\{\Phi_1, \cdots, \Phi_K\}} \mathcal{J}(g) = \inf_{g \in \mathcal{G}_1} \mathcal{J}(g)$$

Proposition

$$\min_{g \in \mathcal{G}_1} \mathcal{L}_1(g) = \min_{\substack{G \in \mathbb{R}^K \\ G^T R G = 1}} G^T H G,$$

where $H := H^{(1)} - H^{(2)} \in \mathbb{R}^{K \times K}$ *is sym psd and*

$$H^{(1)} := \mathbb{E}\left[\|\nabla u(\mathbf{X})\|_2^2 \nabla \Phi(\mathbf{X})^T \nabla \Phi(\mathbf{X}) \right] \in \mathbb{R}^{K \times K},$$

$$H^{(2)} := \mathbb{E}\left[\nabla \Phi(\mathbf{X})^T \nabla u(\mathbf{X}) \nabla u(\mathbf{X})^T \nabla \Phi(\mathbf{X}) \right] \in \mathbb{R}^{K \times K}.$$

1. Introduction
2. Measuring the quality of g
3. Surrogate one feature
4. **Surrogates multiple features**
5. Numerical experiments
6. Conclusion and Perspectives

Surrogates multiple features: Quick overview

- With similar reasoning, we define for $1 \leq j \leq m$,

$$\mathcal{L}_{m,j}(g) := \mathbb{E}\left[\|v_{g,j}(\mathbf{X})\|_2^2 \|\Pi^{\perp}_{v_{g,j}(\mathbf{X})} \Pi^{\perp}_{\nabla g_{-j}(\mathbf{X})} \nabla g_j(\mathbf{X})\|_2^2\right],$$

with $v_{g,j}(\mathbf{x}) := \Pi^{\perp}_{\nabla g_{-j}(\mathbf{x})} \nabla u(\mathbf{x})$.

- For fixed $g_{-j} : \mathbb{R}^d \to \mathbb{R}^{m-1}$, we use $h \mapsto \mathcal{L}_{m,j}((g_{-j}, h))$ as a quadratic surrogate to $h \mapsto \mathcal{J}((g_{-j}, h))$.

- Now need deviation inequalities on $\mathbf{x} \mapsto \|\Pi^{\perp}_{\nabla g_{-j}(\mathbf{x})} \nabla g_j(\mathbf{x})\|_2^2$.

Proposition

Assume the law of $\mathbf{X}$ is s-concave with $s \in (0, 1/d]$. Assume $\mathcal{G}_m$ is a compact set of polynomial with total degree at most $\ell + 1$ with full rank gradients a.s. For a fixed g_{-j}, it holds

$$\gamma_1 \mathcal{L}_{m,j}((g_{-j}, h)) \leq \mathcal{J}((g_{-j}, h)) \leq \gamma_2 \mathcal{L}_{m,j}((g_{-j}, h))^{\frac{1}{1+2\ell m}}$$

Surrogates multiple features: Quick overview

- With similar reasoning, we define for $1 \leq j \leq m$,

$$\mathcal{L}_{m,j}(g) := \mathbb{E}\left[\|v_{g,j}(\mathbf{X})\|_2^2 \|\Pi^{\perp}_{v_{g,j}(\mathbf{X})} \Pi^{\perp}_{\nabla g_{-j}(\mathbf{X})} \nabla g_j(\mathbf{X})\|_2^2\right],$$

with $v_{g,j}(\mathbf{x}) := \Pi^{\perp}_{\nabla g_{-j}(\mathbf{x})} \nabla u(\mathbf{x})$.

- For fixed $g_{-j} : \mathbb{R}^d \to \mathbb{R}^{m-1}$, we use $h \mapsto \mathcal{L}_{m,j}((g_{-j}, h))$ as a quadratic surrogate to $h \mapsto \mathcal{J}((g_{-j}, h))$.

- Now need deviation inequalities on $\mathbf{x} \mapsto \|\Pi^{\perp}_{\nabla g_{-j}(\mathbf{x})} \nabla g_j(\mathbf{x})\|_2^2$.

Proposition

Assume the law of $\mathbf{X}$ is s-concave with $s \in (0, 1/d]$. Assume $\mathcal{G}_m$ is a compact set of polynomial with total degree at most $\ell + 1$ with full rank gradients a.s. For a fixed g_{-j}, it holds

$$\gamma_1 \mathcal{L}_{m,j}((g_{-j}, h)) \leq \mathcal{J}((g_{-j}, h)) \leq \gamma_2 \mathcal{L}_{m,j}((g_{-j}, h))^{\frac{1}{1+2\ell m}}$$

Surrogates multiple features: Quick overview

- With similar reasoning, we define for $1 \leq j \leq m$,

$$\mathcal{L}_{m,j}(g) := \mathbb{E}\left[\|v_{g,j}(\mathbf{X})\|_2^2 \|\Pi^{\perp}_{v_{g,j}(\mathbf{X})} \Pi^{\perp}_{\nabla g_{-j}(\mathbf{X})} \nabla g_j(\mathbf{X})\|_2^2\right],$$

  with $v_{g,j}(\mathbf{x}) := \Pi^{\perp}_{\nabla g_{-j}(\mathbf{x})} \nabla u(\mathbf{x})$.

- For fixed $g_{-j} : \mathbb{R}^d \to \mathbb{R}^{m-1}$, we use $h \mapsto \mathcal{L}_{m,j}((g_{-j}, h))$ as a quadratic surrogate to $h \mapsto \mathcal{J}((g_{-j}, h))$.

- Now need deviation inequalities on $\mathbf{x} \mapsto \|\Pi^{\perp}_{\nabla g_{-j}(\mathbf{x})} \nabla g_j(\mathbf{x})\|_2^2$.

Proposition

Assume the law of $\mathbf{X}$ is s-concave with $s \in (0, 1/d]$. Assume $\mathcal{G}_m$ is a compact set of polynomial with total degree at most $\ell + 1$ with full rank gradients a.s. For a fixed g_{-j}, it holds

$$\gamma_1 \mathcal{L}_{m,j}((g_{-j}, h)) \leq \mathcal{J}((g_{-j}, h)) \leq \gamma_2 \mathcal{L}_{m,j}((g_{-j}, h))^{\frac{1}{1+2\ell m}}$$

Surrogates multiple features: Quick overview

- With similar reasoning, we define for $1 \leq j \leq m$,

$$\mathcal{L}_{m,j}(g) := \mathbb{E}\left[\|v_{g,j}(\mathbf{X})\|_2^2 \|\Pi^{\perp}_{v_{g,j}(\mathbf{X})} \Pi^{\perp}_{\nabla g_{-j}(\mathbf{X})} \nabla g_j(\mathbf{X})\|_2^2\right],$$

with $v_{g,j}(\mathbf{x}) := \Pi^{\perp}_{\nabla g_{-j}(\mathbf{x})} \nabla u(\mathbf{x})$.

- For fixed $g_{-j} : \mathbb{R}^d \to \mathbb{R}^{m-1}$, we use $h \mapsto \mathcal{L}_{m,j}((g_{-j}, h))$ as a quadratic surrogate to $h \mapsto \mathcal{J}((g_{-j}, h))$.

- Now need deviation inequalities on $\mathbf{x} \mapsto \|\Pi^{\perp}_{\nabla g_{-j}(\mathbf{x})} \nabla g_j(\mathbf{x})\|_2^2$.

Proposition

Assume the law of $\mathbf{X}$ is s-concave with $s \in (0, 1/d]$. Assume $\mathcal{G}_m$ is a compact set of polynomial with total degree at most $\ell + 1$ with full rank gradients a.s. For a fixed g_{-j}, it holds

$$\gamma_1 \mathcal{L}_{m,j}((g_{-j}, h)) \leq \mathcal{J}((g_{-j}, h)) \leq \gamma_2 \mathcal{L}_{m,j}((g_{-j}, h))^{\frac{1}{1+2\ell m}}$$

Numerical experiments

Settings

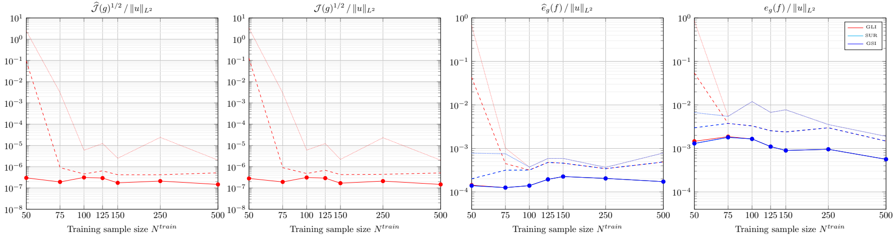
- Feature: Tensorized polynomial basis $\Phi(\mathbf{x}) = (\phi_\alpha(\mathbf{x}))_{\alpha \in \Lambda_{p,k}}$ with

$$\phi_\alpha(\mathbf{x}) := \prod_{\nu=1}^{d} \phi^{\nu}_{\alpha_\nu}(x_\nu), \quad \Lambda_{p,k} := \{\alpha \in \mathbb{N}^d : \|\alpha\|_p \leq k\} \setminus \{0\},$$

  where (k, p) are hyperparameters learnt by 5-fold cross-validation.

- Regression: Kernel regression with gaussian kernel and Ridge regularization, whose hyperparameters are learnt by 10-fold cross-validation.

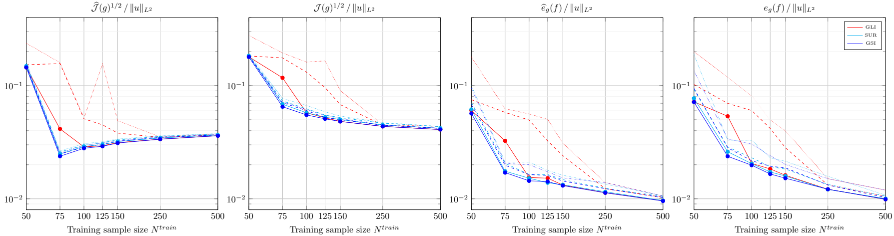
Numerical experiments: one feature



$$u(x) = \sin(\frac{4}{\pi^2}\|x\|^2), \quad d = 8, \quad m = 1$$

50%, 90% and 100% quantiles over 20 realizations.
Left plots : train and test estimations of $\mathcal{J}(g)/\|u\|_{L^2}$.
Right plots : train and test estimations of $\|u - f \circ g\|_{L^2}/\|u\|_{L^2}$.

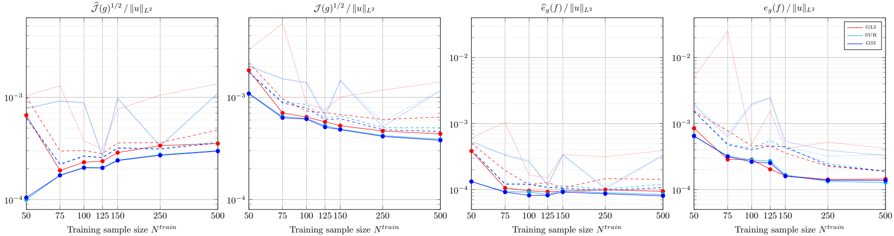
Numerical experiments: one feature



$$u(x) = \exp\left(\frac{1}{d}\sum_{i=1}^{d} \sin(x_i) e^{\cos(x_i)}\right), \quad \mathbf{X} \sim \mathcal{U}(]-\frac{\pi}{2}, \frac{\pi}{2}[^8) \quad m = 1$$

50%, 90% and 100% quantiles over 20 realizations.
Left plots : train and test estimations of $\mathcal{J}(g)/\|u\|_{L^2}$.
Right plots : train and test estimations of $\|u - f \circ g\|_{L^2}/\|u\|_{L^2}$.

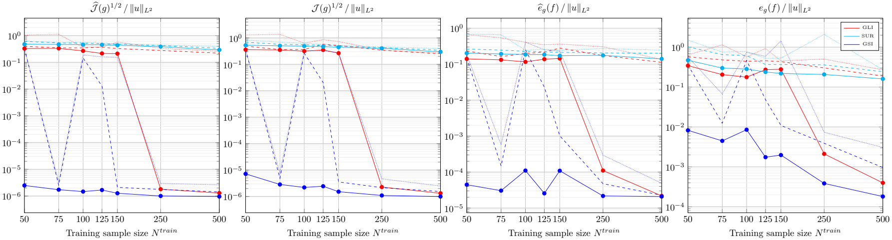
Numerical experiments: one feature



$$u(x) = \text{borehole}, \quad d = 8, \quad m = 1$$

50%, 90% and 100% quantiles over 20 realizations.
Left plots : train and test estimations of $\mathcal{J}(g)/\|u\|_{L^2}$.
Right plots : train and test estimations of $\|u - f \circ g\|_{L^2}/\|u\|_{L^2}$.

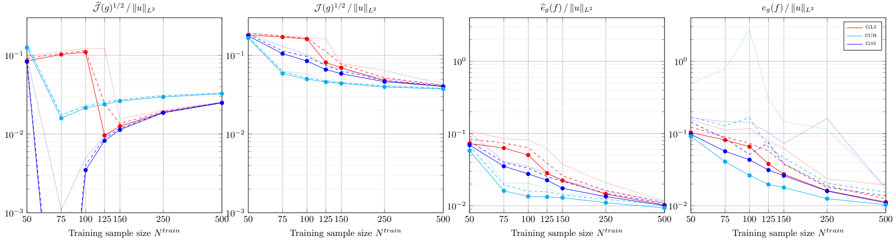
Numerical experiments: multiple features



$$u(x) := \cos(\frac{1}{2}x^T x) + \sin(\frac{1}{2}x^T M x), \quad \mathbf{X} \sim \mathcal{U}(]-\frac{\pi}{2}, \frac{\pi}{2}[^8) \quad m = 2$$

50%, 90% and 100% quantiles over 20 realizations.
Left plots : train and test estimations of $\mathcal{J}(g)/\|u\|_{L^2}$.
Right plots : train and test estimations of $\|u - f \circ g\|_{L^2}/\|u\|_{L^2}$.

Numerical experiments: multiple features



$$u(x) = \exp\left(\frac{1}{d}\sum_{i=1}^{d} \sin(x_i) e^{\cos(x_i)}\right), \quad \mathbf{X} \sim \mathcal{U}(]-\frac{\pi}{2}, \frac{\pi}{2}[^8) \quad m = 2$$

50%, 90% and 100% quantiles over 20 realizations.
Left plots : train and test estimations of $\mathcal{J}(g)/\|u\|_{L^2}$.
Right plots : train and test estimations of $\|u - f \circ g\|_{L^2}/\|u\|_{L^2}$.

Conclusion and Perspectives

- Quadratic surrogates with controlled suboptimality
- Works well for $m = 1$, more mitigated for $m > 1$.

→ Structured approach in a Tensor-Network fashion.

→ Optimal sampling for g.

→ Optimal sampling for f, oblivious to g.

Conclusion and Perspectives

- Quadratic surrogates with controlled suboptimality
- Works well for $m = 1$, more mitigated for $m > 1$.

$\rightarrow$ Structured approach in a Tensor-Network fashion.

$\rightarrow$ Optimal sampling for g.

$\rightarrow$ Optimal sampling for f, oblivious to g.

Thank you !

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.

Fradelizi, M. (2009).
Concentration inequalities for s-concave measures of dilations of Borel sets and applications.
Electron. J. Probab., 14(none).

References II

Romor, F., Tezzele, M., Lario, A., and Rozza, G. (2022).
Kernel-based active subspaces with application to computational fluid dynamics parametric problems using the discontinuous Galerkin method.
Numerical Meth Engineering, 123(23):6000–6027.

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

Zhang, G., Zhang, J., and Hinkle, J. (2019).
Learning nonlinear level sets for dimensionality reduction in function approximation.
In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.