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

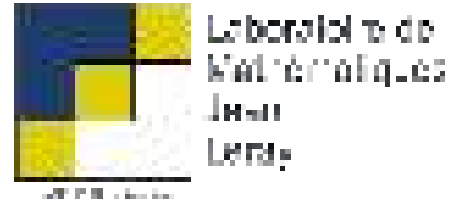
Alexandre PASCO [1,(2)] Anthony NOUY [1]

[1]LMJL, Centrale Nantes, Nantes Université, France
[(2)]Chair HPNALGS, EPFL, Switzerland

Curves and Surfaces, June 2026



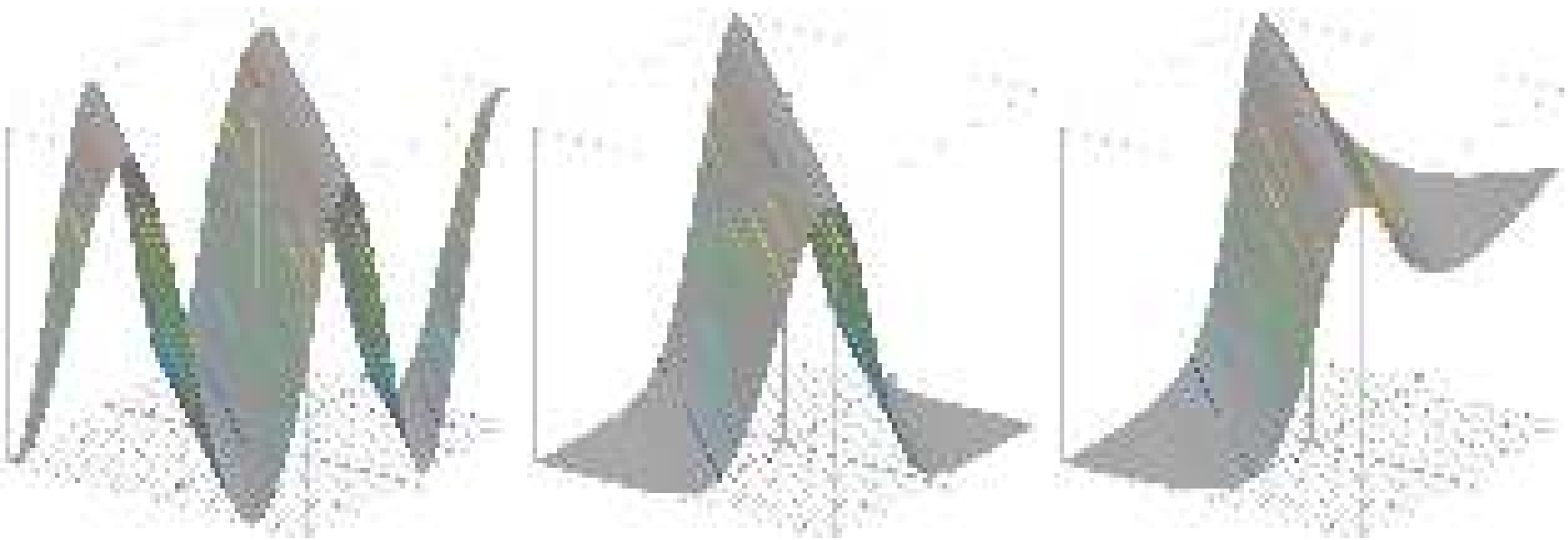




1. Introduction
2. Gradient-based dimension reduction
3. Quadratic surrogate to the loss
4. Numerical experiments
5. Conclusion

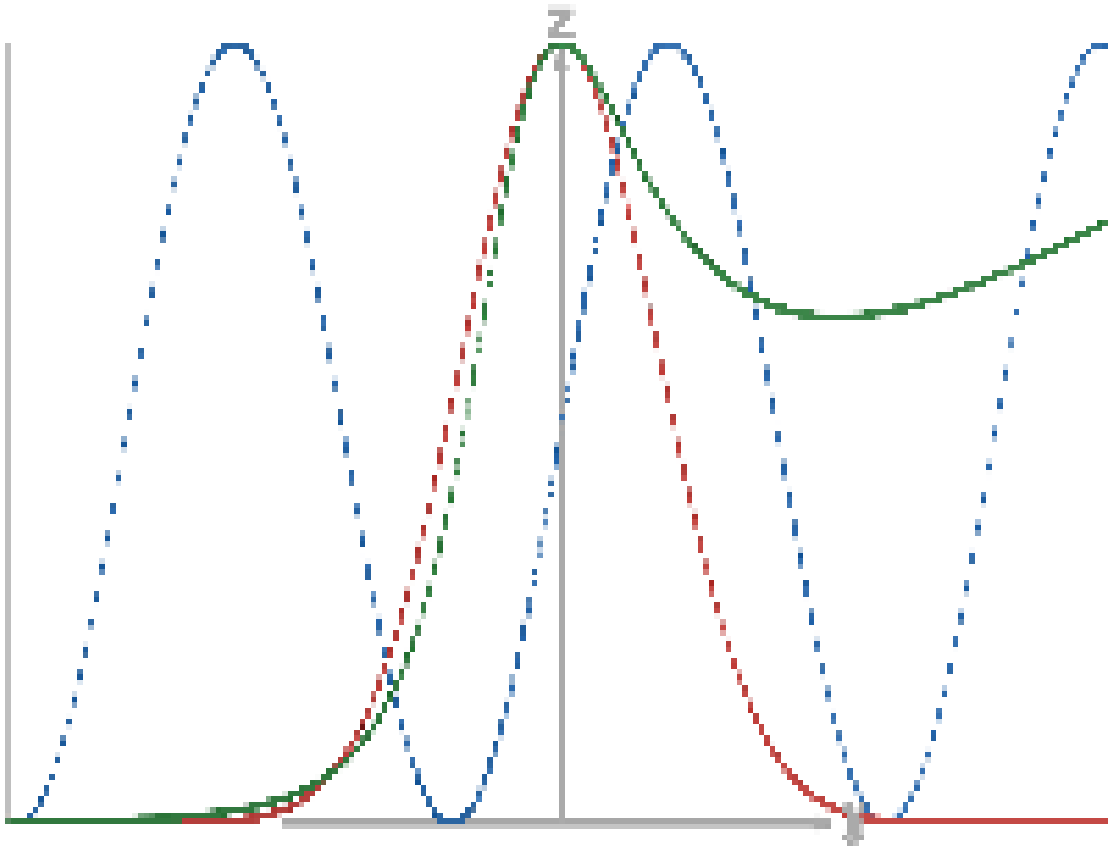
Introduction: collective dimension reduction

$u(\mathbf{x}, 1)$ $u(\mathbf{x}, 2)$ $u(\mathbf{x}, 2)$



Introduction: collective dimension reduction

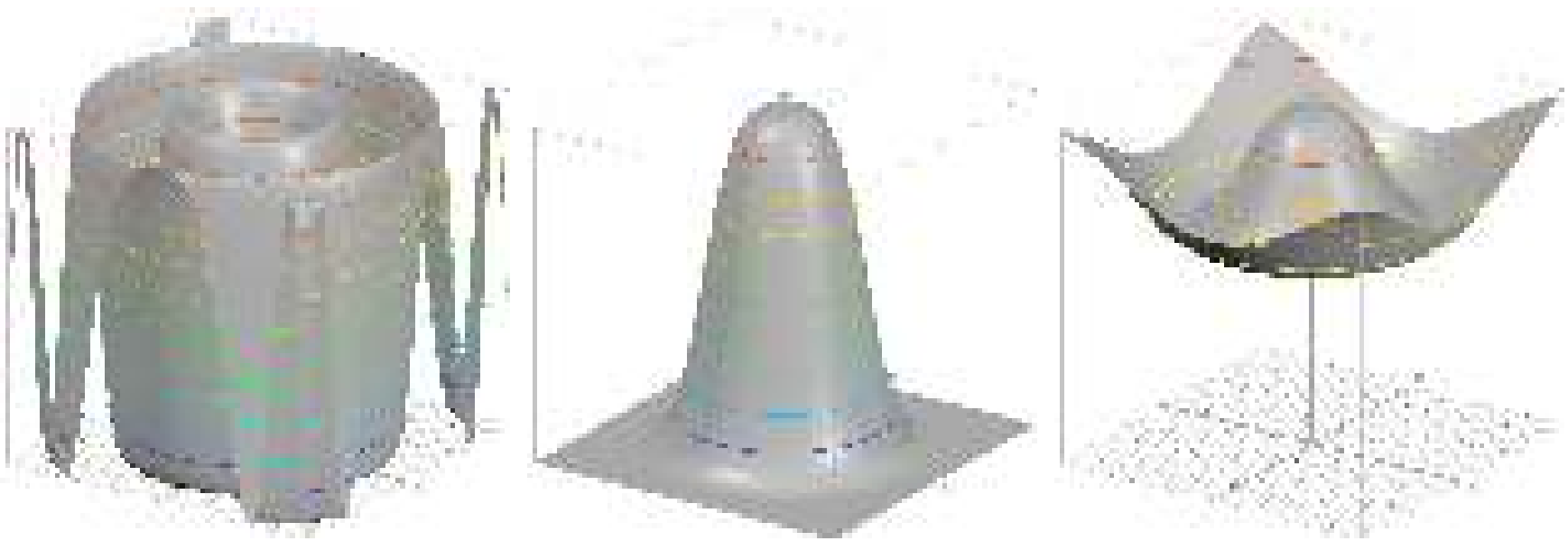
$u(\mathbf{x}, 1)$ $u(\mathbf{x}, 2)$ $u(\mathbf{x}, 2)$



$g(\mathbf{x}) = x_1 + x_2$

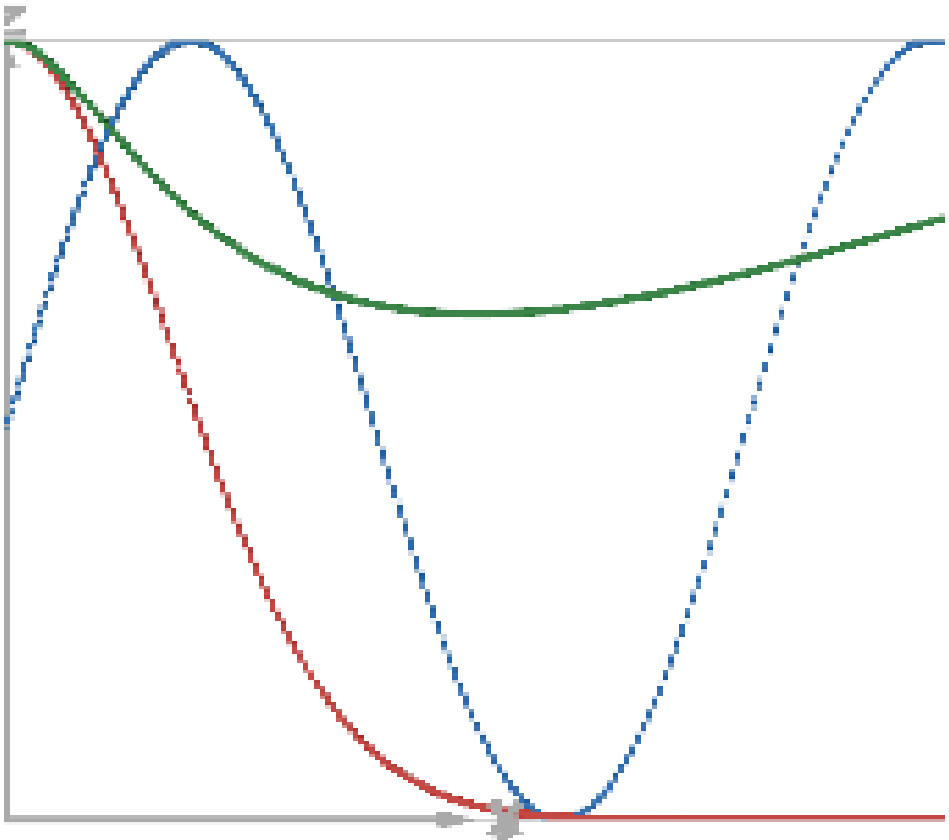
Introduction: collective dimension reduction

$u(\mathbf{x}, 1)$ $u(\mathbf{x}, 2)$ $u(\mathbf{x}, 2)$



Introduction: collective dimension reduction

$u(\mathbf{x}, 1)$ $u(\mathbf{x}, 2)$ $u(\mathbf{x}, 2)$



$$g(\mathbf{x}) = \|\mathbf{x}\|_2^2$$

Introduction: formal setting

- Goal: approximate $u(\mathbf{X}, Y) \in \mathbb{R}$, random $\mathbf{X} \in \mathbb{R}^d$ and $Y \in \mathcal{Y}$.
  $d \gg 1$ requires structured approximation (low-rank, sparse, ...).

- Approximation format:

$$\hat{u}(\mathbf{X}, Y) = f(g(\mathbf{X}), Y), \quad g(\mathbf{X}) \in \mathbb{R}^m, \quad m \leq d.$$

  Use the same features $g(\mathbf{X})$ for every elements of $\{u(\mathbf{X}, y) : y \in \mathcal{Y}\}$.

- Today's scope: learn the feature map $g : \mathbb{R}^d \to \mathbb{R}^m$, forget about f.

- Linear features: ridge function (Logan & Shepp, 1975; Pinkus, 2015),

$$\hat{u}(\mathbf{X}) = f(G^T \mathbf{X}, Y),$$

Introduction: examples

Can be seen as dimension reduction on $v(\mathbf{X}) \in \mathcal{V}$.

- $\mathcal{Y} = \{y_0\}$ then non-collective,

$$v(\mathbf{X}) = u(\mathbf{X}, y_0).$$

- $Y \sim \mathcal{U}(\{y_1, \cdots, y_n\})$ then $\mathcal{V} = \mathbb{R}^n$ and

$$v(\mathbf{X}) = (u(\mathbf{X}, y_1), \cdots, u(\mathbf{X}, y_n)).$$

- $Y \sim \mathcal{U}(0, 1)$ then $\mathcal{V} = L^2(0, 1)$ and

$$v(\mathbf{X}) = u(\mathbf{X}, \cdot).$$

Big picture preview

- $\mathcal{J}(g)$ as upper bound of the best approximation error of $u(\mathbf{X}, Y)$ by a function of $g(\mathbf{X})$ and Y.

- Split $\mathcal{J}(g) \sim \mathcal{J}_m(g) + \varepsilon_m$ and focus on $\mathcal{J}_m$.

- Quadratic surrogate $\mathcal{L}_m$ that achieves exact recovery. Also for g polynomial of degree $\ell + 1$ and some probability distributions,

$$\mathcal{J}_m(g) \lesssim \mathcal{L}_m(g)^{1/(1+2\ell m)}.$$

1 Introduction

2 Gradient-based dimension reduction

3 Quadratic surrogate to the loss

4 Numerical experiments

5 Conclusion

Informally: align the gradients

- Non-collective: if $u(\mathbf{x}) = f(g(\mathbf{x}))$, the chain rule yields

$$\nabla u(\mathbf{x}) = \nabla g(\mathbf{x}) \nabla f(g(\mathbf{x})),$$

which implies

$$\nabla u(\mathbf{x}) \in \operatorname{span}\{\nabla g(\mathbf{x})\} := \operatorname{span}\{\nabla g_1(\mathbf{x}), \cdots, \nabla g_m(\mathbf{x})\} \subset \mathbb{R}^d.$$

$\rightarrow$ Maximize the alinement between $\nabla g(\mathbf{x})$ and $\nabla u(\mathbf{x})$.

- Collective: if $u(\mathbf{x}, y) = f(g(\mathbf{x}), y)$, with the same reasoning,

$$\nabla_{\mathbf{x}} u(\mathbf{x}, y) \in \operatorname{span}\{\nabla g(\mathbf{x})\}.$$

$\rightarrow$ Maximize the alinement between $\nabla g(\mathbf{x})$ and $\{\nabla_{\mathbf{x}} u(\mathbf{x}, y) : y \in \mathcal{Y}\}$

Formally: align the gradients

- Non-collective: (Bigoni et al., 2022)

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

  with $C_{\mathbf{X}|\mathcal{G}_m}$ defined through Poincaré inequalities on manifolds, and

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

  Challenging to minimize on nonlinear features due to $\Pi_{\nabla g(\mathbf{X})}$.

- Collective: apply the above to each elements $\{u(\cdot, y) : y \in \mathcal{Y}\}$

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

- Today's scope: focus on $\mathcal{J}(g)$, forget about $C_{\mathbf{X}|\mathcal{G}_m}$.
  See (Verdière et al., 2025; Zahm et al., 2020), also (Pillaud-Vivien et al., 2020).

Aligning gradients: drawing

Non-collective and linear features $\rightarrow$ PCA of $\nabla u(\mathbf{X})$.



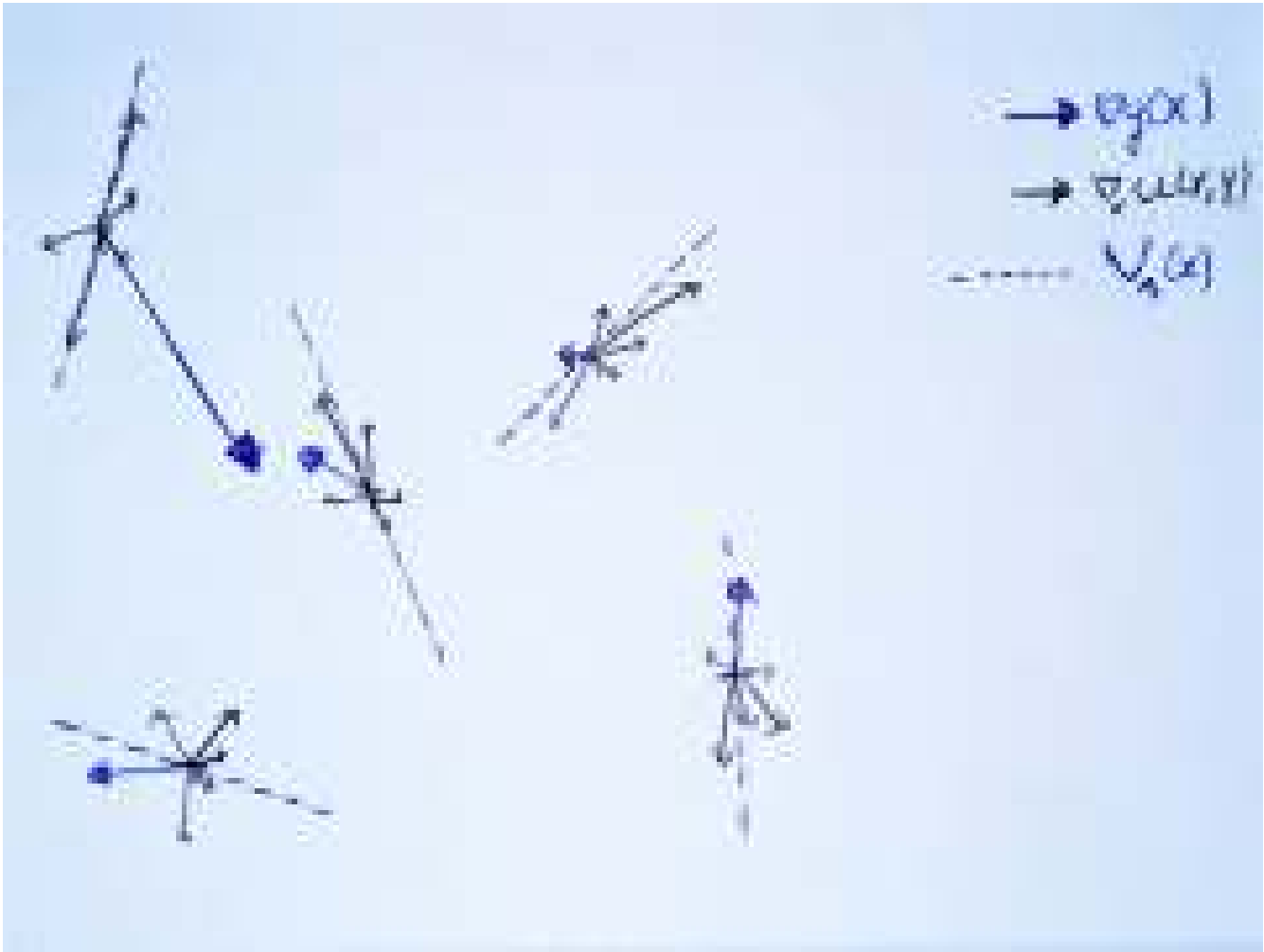
Aligning gradients: drawing

Non-collective and nonlinear features $\rightarrow$ Hard problem.



Aligning gradients: drawing

Collective and nonlinear features $\rightarrow$ Hard, cannot beat local PCA.



Aligning gradients: drawing

Collective and nonlinear features conditioned to $\mathbf{X}$.



Eigenmodes truncation as benchmark

- Fixe $\mathbf{x} \in \mathbb{R}^d$, we search a subspace spanned by $V(\mathbf{x}) \in \mathbb{R}^{d \times m}$ approximating the parameterized set $\{\nabla_{\mathbf{x}} u(\mathbf{x}, y) : y \in \mathcal{Y}\} \subset \mathbb{R}^d$.
- Best choice: $V_m(\mathbf{x})$ as the dominant eigenmodes of

$$M(\mathbf{x}) := \mathbb{E}_Y \left[\nabla_{\mathbf{x}} u(\mathbf{x}, Y) \nabla_{\mathbf{x}} u(\mathbf{x}, Y)^T\right] \in \mathbb{R}^{d \times d},$$

  with optimal error $\varepsilon_m := \sum_{i=m+1}^{d} \mathbb{E}\left[\lambda_i(M(\mathbf{X}))\right]$ such that

$$\mathbb{E}\left[\|\nabla_{\mathbf{x}} u(\mathbf{X}, Y)\|_2^2 - \|\Pi_{V(\mathbf{X})} \nabla_{\mathbf{x}} u(\mathbf{X}, Y)\|_2^2\right] = \varepsilon_m.$$

- "Constraint": subspaces of the form $\operatorname{span}\{V(\mathbf{x})\} = \operatorname{span}\{\nabla g(\mathbf{x})\}$,

$$\mathbb{E}\left[\|\nabla_{\mathbf{x}} u(\mathbf{X}, Y)\|_2^2 - \|\Pi_{\nabla g(\mathbf{X})} \nabla_{\mathbf{x}} u(\mathbf{X}, Y)\|_2^2\right] = \mathcal{J}(g) \geq \varepsilon_m.$$

New loss function from truncated gradient

- Initially: align $\nabla g(\mathbf{x})$ and $\{\nabla_{\mathbf{x}} u(\mathbf{x}, y) : y \in \mathcal{Y}\}$.
- Now: align $\nabla g(\mathbf{x})$ and $\{\Pi_{V_m(\mathbf{x})} \nabla_{\mathbf{x}} u(\mathbf{x}, y) : y \in \mathcal{Y}\}$.
- New truncated loss function, with $\Pi^{\perp}_{\nabla g(\mathbf{x})} = I_d - \Pi_{\nabla g(\mathbf{x})}$,

$$\mathcal{J}_m(g) := \mathbb{E}\left[\|\Pi^{\perp}_{\nabla g(\mathbf{X})} \Pi_{V_m(\mathbf{X})} \nabla_{\mathbf{x}} u(\mathbf{X}, Y)\|_2^2\right].$$

Proposition (P&N, 2026)

For all $g \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^m)$,

$$\frac{1}{2}(\mathcal{J}_m(g) + \varepsilon_m) \leq \mathcal{J}(g) \leq \mathcal{J}_m(g) + \varepsilon_m.$$

Proof: $V_m(\mathbf{X})$ spans an eigenspace, projectors have operator norm 1.

Why bother ?

$$\mathcal{J}(g) = \mathbb{E}\left[\|\Pi^{\perp}_{\nabla g(\mathbf{X})} \nabla_{\mathbf{x}} u(\mathbf{X}, Y)\|_2^2\right],$$
$$\mathcal{J}_m(g) = \mathbb{E}\left[\|\Pi^{\perp}_{\nabla g(\mathbf{X})} \Pi_{V_m(\mathbf{X})} \nabla_{\mathbf{x}} u(\mathbf{X}, Y)\|_2^2\right].$$

Two major problems.

- Minimizing $\mathcal{J}_m$ is at least as challenging as for $\mathcal{J}$.
- Estimating ε_m, $M(\mathbf{X})$ and $V_m(\mathbf{X})$ requires specific sampling.

Then why bother ?

+ $\mathcal{J}_m$ is better suited to quadratic surrogates as in (N&P 2026).
+ ε_m is a benchmark error independent of $\mathcal{G}_m$.

1 Introduction

2 Gradient-based dimension reduction

3 Quadratic surrogate to the loss

4 Numerical experiments

5 Conclusion

Towards the quadratic surrogate

- Intermediate result (P&N, 2026): if $\nabla g(\mathbf{X})$ is full rank a.s., then

$$\mathbb{E}\left[\frac{\lambda_m(M(\mathbf{X}))}{\sigma_1(\nabla g(\mathbf{X}))^2}\underbrace{\|\Pi^{\perp}_{V_m(\mathbf{X})}\nabla g(\mathbf{X})\|_F^2}_{\text{quadratic in } g}\right] \leq \mathcal{J}_m(g)$$

$$\leq \mathbb{E}\left[\frac{\lambda_1(M(\mathbf{X}))}{\sigma_m(\nabla g(\mathbf{X}))^2}\underbrace{\|\Pi^{\perp}_{V_m(\mathbf{X})}\nabla g(\mathbf{X})\|_F^2}_{\text{quadratic in } g}\right].$$

- We want an surrogate as an upper-bound, we thus define

$$\mathcal{L}_m(g) := \mathbb{E}\left[\lambda_1(M(\mathbf{X}))\|\Pi^{\perp}_{V_m(\mathbf{X})}\nabla g(\mathbf{X})\|_F^2\right].$$

This surrogate is quadratic, thus better suited to optimization.

Surrogate: exact recovery

$$\mathcal{J}_m(g) = \mathbb{E}\left[\|\Pi^{\perp}_{\nabla g(\mathbf{X})}\Pi_{V_m(\mathbf{X})}\nabla_{\mathbf{x}}u(\mathbf{X},Y)\|_2^2\right],$$

$$\mathcal{L}_m(g) = \mathbb{E}\left[\lambda_1(M(\mathbf{X}))\|\Pi^{\perp}_{V_m(\mathbf{X})}\nabla g(\mathbf{X})\|_F^2\right],$$

$V_m(\mathbf{x})$ first eigenmodes of $M(\mathbf{x}) = \mathbb{E}_Y\left[\nabla u_{\mathbf{x}}(\mathbf{x},Y)\nabla u_{\mathbf{x}}(\mathbf{x},Y)^T\right]$.

Proposition (Exact recovery (P&N, 2026))

If $\mathrm{rank}(M(\mathbf{X})) \geq \mathrm{rank}(\nabla g(\mathbf{X})) = m$ *a.s.*,

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

Proof: $\|\Pi^{\perp}_{V_m(\mathbf{X})}\nabla g(\mathbf{X})\|_F^2 = 0$ iff $\mathrm{span}\{\nabla g(\mathbf{X})\} \subset \mathrm{span}\{V_m(\mathbf{X})\}$.
A.s., both spaces have dim. m, thus they are equal and
$\Pi^{\perp}_{\nabla g(\mathbf{X})}\Pi_{V_m(\mathbf{X})} = 0$ a.s.

Surrogate: upper bound

$$\mathcal{J}_m(g) = \mathbb{E}\left[\|\Pi^{\perp}_{\nabla g(\mathbf{X})}\Pi_{V_m(\mathbf{X})}\nabla_{\mathbf{x}} u(\mathbf{X}, Y)\|_2^2\right],$$

$$\mathcal{L}_m(g) = \mathbb{E}\left[\lambda_1(M(\mathbf{X}))\|\Pi^{\perp}_{V_m(\mathbf{X})}\nabla g(\mathbf{X})\|_F^2\right],$$

$V_m(\mathbf{x})$ first eigenmodes of $M(\mathbf{x}) = \mathbb{E}_Y\left[\nabla u_{\mathbf{x}}(\mathbf{x}, Y)\nabla u_{\mathbf{x}}(\mathbf{x}, Y)^T\right]$.

Proposition (Polynomial features (P&N, 2026))

If $\operatorname{rank}(M(\mathbf{X})) \geq m$ *a.s., if* $\mathbf{X} \sim \mathcal{U}(\mathcal{X})$ *with* $\mathcal{X}$ *convex, if* $\mathcal{G}_m$ *is a closed set of poly. of total degree at most* $\ell + 1$ *whose gradient rank is full a.s, then*

$$\mathcal{J}_m(g) \lesssim \mathcal{L}_m(g)^{\frac{1}{1+2\ell m}}.$$

Proof: Step 1, condition intermediate result to $\{\sigma_m(\nabla g(\mathbf{X}))^2 < \varepsilon\}$.
Step 2, poly. anti-concentration $\mathbb{P}\left[\sigma_m(\nabla g(\mathbf{X}))^2 < \varepsilon\right] \lesssim \varepsilon^{1/2\ell m}$.
Step 3, optimize resulting upper bound over ε.

Surrogate: lower bound only non-collective and $m = 1$

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

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

Proposition (N&P, 2026)

If $\mathbf{X} \sim \mathcal{U}(\mathcal{X})$ is compactly supported and $\|\nabla g(\mathbf{X})\|_2^2 \leq C$ a.s, then

$$\mathcal{J}(g) \gtrsim \mathcal{L}(g)$$

Proof: Step 1, both sides in the intermediate result are equal.
Step 2, substitute uniform bound in intermediate result.

(N&P, 2026) More general results for non-collective:

- $m = 1$, s-concave proba measures on $\mathbf{X}$ (including log-concave).
- $m > 1$, bounds when fixing all but one feature.

Surrogate as minimum eigenvalue problem

- Fixe basis $\Phi : \mathbb{R}^d \to \mathbb{R}^K$, $K \geq d$, of nonlinear features.

- For all $G \in \mathbb{R}^{K \times m}$, $\mathcal{L}_m(G^T\Phi) = \mathrm{Tr}\left(G^T H_m G\right)$, with

$$H_m = \mathbb{E}\left[\lambda_1(M(\mathbf{X}))\nabla\Phi(\mathbf{X})(I_d - V_m(\mathbf{X})V_m(\mathbf{X})^T)\nabla\Phi(\mathbf{X})^T\right] \in \mathbb{R}^{K\times K}.$$

- If Φ is polynomial and $\mathbb{E}\left[\nabla\Phi(\mathbf{X})^T\nabla\Phi(\mathbf{X})\right] = I_K$, then assumptions on $\mathcal{G}_m$ for previous results are satisfied, thus

$$\min_G \mathcal{J}_m(G^T\Phi) = \min_{\substack{G, \\ G^TG=I_m}} \mathcal{J}_m(G^T\Phi) \quad \text{(invariance property)}$$

$$\lesssim \left(\min_{\substack{G, \\ G^TG=I_m}} \mathrm{Tr}\left(G^T H_m G\right)\right)^{\frac{1}{1+2\ell m}}.$$

In practice

- Estimate expectations from samples.
  $\rightarrow$ Surrogate for collective: tensorized samples.
  $\rightarrow$ Surrogate for non-collective: arbitrary samples.

- Minimizer of $\mathcal{L}_m$ as initialization for iterative minimization of $\mathcal{J}$.

- Cross validation procedure for both $g \in \mathcal{G}_m$ and $f \in \mathcal{F}_m$.

- Implementation with tutorials
  https://github.com/anthony-nouy/tensap

1 Introduction

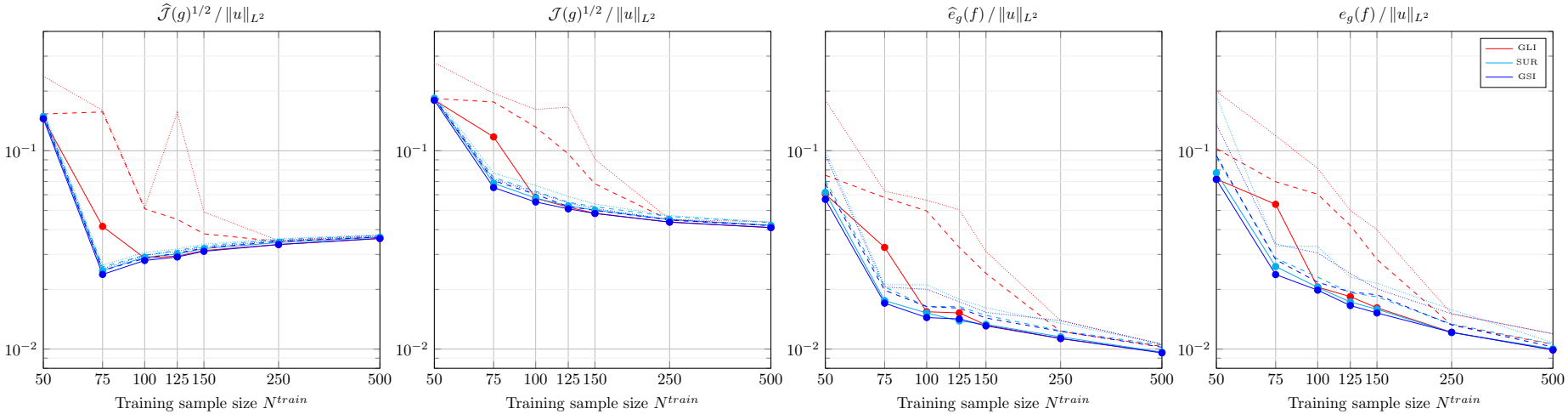
2 Gradient-based dimension reduction

3 Quadratic surrogate to the loss

4 Numerical experiments

5 Conclusion

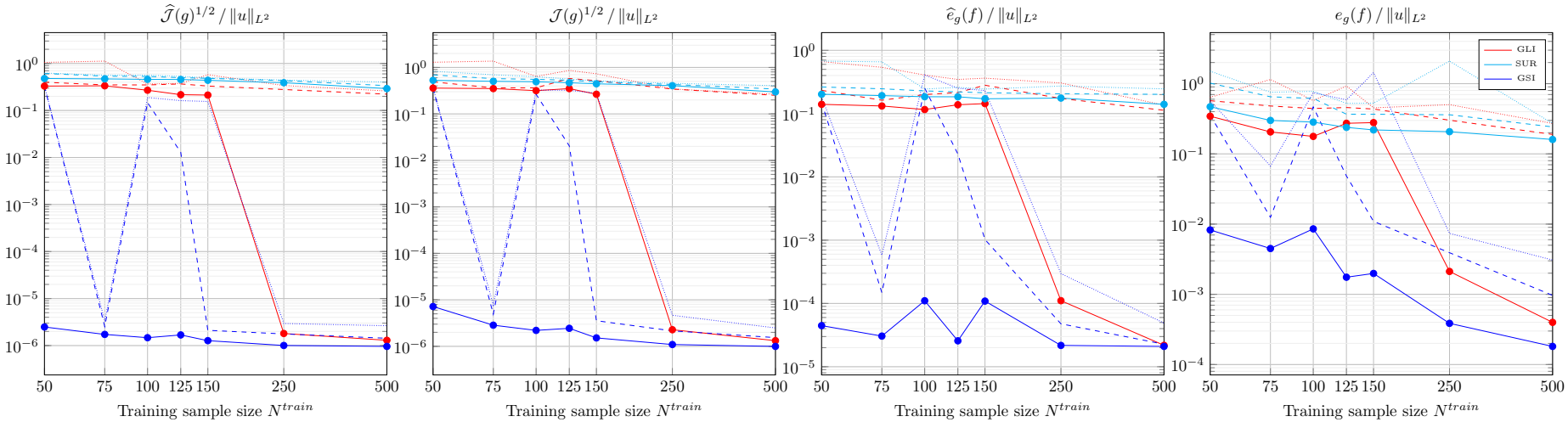
Numerical experiment: non-collective approx $m = 1$



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

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 $\|u - f \circ g\|_{L^2}/\|u\|_{L^2}$.
Red: quasi-Newton minimization of $\mathcal{J}$.
Cyan: minimization of the surrogate.
Blue: surrogate as initialization of quasi-Newton for $\mathcal{J}$.

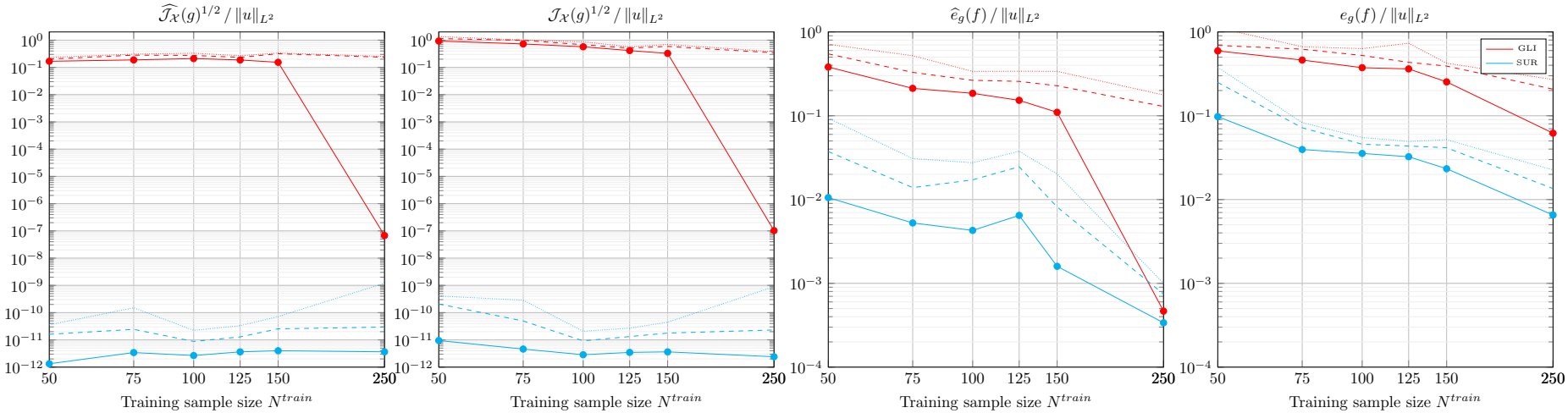
Numerical experiment: non-collective exact $m = 2$



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

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 $\|u - f \circ g\|_{L^2}/\|u\|_{L^2}$.
Red: quasi-Newton minimization of $\mathcal{J}$.
Cyan: minimization of the surrogate.
Blue: surrogate as initialization of quasi-Newton for $\mathcal{J}$.

Numerical experiment: collective exact $m = 3$



$$u(\mathbf{x}, y) := \sum_{k=1}^{p} (\mathbf{x}^T Q_k \mathbf{x})^2 \sin(\frac{\pi k}{2p} y), \quad \mathbf{X} \sim \mathcal{U}(]-1,1[^8), Y \sim \mathcal{U}(]-1,1[), p = m = 3.$$

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 $\|u - f \circ g\|_{L^2}/\|u\|_{L^2}$.
Red: quasi-Newton minimization of $\mathcal{J}$.
Cyan: minimization of the surrogate.

1 Introduction

2 Gradient-based dimension reduction

3 Quadratic surrogate to the loss

4 Numerical experiments

5 Conclusion

Big picture

- $\mathcal{J}_{\mathcal{X}}(g)$ as upper bound of the best approximation error of $u(\mathbf{X}, Y)$ by a function of $g(\mathbf{X})$ and Y.
  $\rightarrow$ $\mathcal{J}_{\mathcal{X}}(g)$ measures the alignment between $\nabla g(\mathbf{X})$ and $\nabla_{\mathbf{x}} u(\mathbf{X}, Y)$.
  $\rightarrow$ Direct extension of the P.I.-based approach (Bigoni et al., 2022).

- Split $\mathcal{J}_{\mathcal{X}}(g) \sim \mathcal{J}_m(g) + \varepsilon_m$ and focus on $\mathcal{J}_m$.
  $\rightarrow$ Benchmark ε_m from $\mathbb{E}_{\mathbf{X}}$ of PCA/POD approx. of $y \mapsto \nabla_{\mathbf{x}} u(\mathbf{x}, y)$.
  $\rightarrow$ Truncated loss $\mathcal{J}_m(g)$ measures the alignment between $\nabla g(\mathbf{X})$ and dominant PCA/POD modes $V_m(\mathbf{X})$.

- Quadratic surrogate $\mathcal{L}_m$ that achieves exact recovery. Also $\mathcal{J}_m(g) \lesssim \mathcal{L}_m(g)^{1/(1+2\ell m)}$ for poly. g and some proba. distrib.
  $\rightarrow$ More general and converse bounds available for $m = 1$.
  /!\ Collective setting requires tensorized samples $(\mathbf{x}^{(i)}, y^{(j)})$.

Take home messages

Take home messages

- Gradient-based dimension reduction is competitive when few samples are available and good dimension reduction possible ($m < d$).
- Surrogate is competitive when drastic reduction is possible ($m \ll d$).
- Disclaimer: no rigorous error certification for general nonlinear features.

Perspective and open question:

- Compression of "fast-to-evaluate" functions u.
- Collective approximation error as $m \to d$? "Feature width" ?

Thank you for your attention.

References I

Bigoni, D., Marzouk, Y., Prieur, C., & Zahm, O. (2022). Nonlinear dimension reduction for surrogate modeling using gradient information. *Information and Inference: A Journal of the IMA*, *11*(4), 1597–1639. https://doi.org/10.1093/imaiai/iaac006

Logan, B. F., & Shepp, L. A. (1975). Optimal reconstruction of a function from its projections. *Duke Mathematical Journal*, *42*(4). https://doi.org/10.1215/S0012-7094-75-04256-8

Nouy, A., & Pasco, A. (2026). Surrogate to Poincaré inequalities on manifolds for dimension reduction in nonlinear feature spaces. *Information and Inference: A Journal of the IMA*, *15*(1), iaag005. https://doi.org/10.1093/imaiai/iaag005

Pasco, A., & Nouy, A. (2026). Surrogate to Poincaré inequalities on manifolds for structured dimension reduction in nonlinear feature spaces. https://doi.org/10.48550/ARXIV.2602.01143

References II

Pillaud-Vivien, L., Bach, F., Lelièvre, T., Rudi, A., & Stoltz, G. (2020). Statistical estimation of the poincaré constant and application to sampling multimodal distributions. In S. Chiappa & R. Calandra (Eds.), *Proceedings of the twenty third international conference on artificial intelligence and statistics* (pp. 2753–2763, Vol. 108). PMLR.

Pinkus, A. (2015, August). *Ridge Functions* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781316408124

Verdière, R., Prieur, C., & Zahm, O. (2025). Diffeomorphism-based feature learning using Poincaré inequalities on augmented input space. *Journal of Machine Learning Research*, *26*(139), 1–31. https://doi.org/hal-04364208

References III

Zahm, O., Constantine, P. G., Prieur, C., & Marzouk, Y. M. (2020). Gradient-Based Dimension Reduction of Multivariate Vector-Valued Functions. *SIAM Journal on Scientific Computing*, *42*(1), A534–A558. https://doi.org/10.1137/18M1221837

Towards structured features

For multi-index $\alpha \subset \{1, \cdots, d\}$, consider the approximation with structured features

$$\hat{u}(\mathbf{X}) = f(g^{\alpha}(\mathbf{X}_{\alpha}), g^{\alpha^c}(\mathbf{X}_{\alpha^c})).$$

Link with collective problems (P&N, 2026):

- The optimal features g^{α} and g^{α^c} from collectives problems (taking resp. $Y = \mathbf{X}_{\alpha^c}$ and $Y = \mathbf{X}_{\alpha}$) are nearly optimal wrt L^2 error. Also holds to N independent groups of variables.
  $\rightarrow$ Similar to HOSVD results for Tucker-tensor format.

- They are exactly optimal wrt the P.I.-based loss $\mathcal{J}$.

On the constant in the upper-bound

Manifold valued $\mathbf{X}' \in \mathcal{M}$, define $C_{\mathbf{X}'}$ the Poincaré constant associated to $\mathbf{X}'$ as the smallest C satisfying

$$\forall h \in \mathcal{C}^1(\mathcal{M}), \quad \mathbb{V}\left[h(\mathbf{X}')\right] \leq C\mathbb{E}\left[\|\nabla h(\mathbf{X}')\|^2\right].$$

Consider the conditional random variables $\mathbf{X}|h(\mathbf{X}) = \mathbf{z}$,

$$C_{\mathbf{X}|\mathcal{G}_m} := \sup_{h \in \mathcal{G}_m} \sup_{\mathbf{z} \in \mathbb{R}^m} C_{\mathbf{X}|h(\mathbf{X})=\mathbf{z}}.$$

Ensuring $C_{\mathbf{X}|\mathcal{G}_m} < +\infty$ is challenging for nonlinear features.
Heuristic approaches to minimize it:

- (Verdière et al., 2025) Penalize by $D_{KL}(\psi(\mathbf{X})||\mathbf{Z}')$ with $\mathbf{Z}' \sim \mathcal{N}(0, I_m)$, since $C_{\mathbf{Z}'|\mathbf{Z}_1,\cdots,\mathbf{Z}_m=\mathbf{z}} = 1$.

- (Pillaud-Vivien et al., 2020) Maximize sample-based estimation of $C_{h(\mathbf{X})}$ so that $h(\mathbf{X})$ enraptures most of multimodality.

On the objective function

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

- Linear features: $\mathbf{x} \mapsto \Pi_{\nabla g(\mathbf{X})}$ is constant, $G \mapsto \mathcal{J}(G^T)$ is a Rayleigh quotient and is minimized by the **dominant eigenvectors** of

$$\mathbb{E}\left[\nabla u(\mathbf{X})\nabla u(\mathbf{X})^T\right] \in \mathbb{R}^{d\times d}.$$

- Nonlinear features: $g(\mathbf{x}) = G^T\Phi(\mathbf{x})$ 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.

Approximation properties

- We are approximating the set $\mathcal{M}_u := \{\mathbf{x} \mapsto u(\mathbf{x}, y), \quad y \in \mathcal{Y}\}$ by an (infinite-dimensional) subspace

$$\mathcal{V}_g := \{(\mathbf{x}, y) \mapsto f(g(\mathbf{x}), y), \ f \in L^2(g_{\#}\mu_{\mathbf{X}} \otimes \mu_Y)\}.$$

  Question: how fast is the decay of

$$e_m(u) := \inf_{\substack{g:\mathcal{X}\to\mathbb{R}^m, \\ \text{regularity}(g)}} \text{dist}(\mathcal{M}_u, \mathcal{V}_g) \quad \text{as} \quad m \to \dim(\mathcal{X}) = d \quad ?$$

- (Pinkus, 2015) Approximation properties mainly available for linear g.

- (Pinkus, 2015, Example 7.3) In general $\mathcal{V}_g + \mathcal{V}_h$ is not closed in $L^2(\mu_{\mathbf{X}} \otimes \mu_Y)$ because Y is a shared direction.