Nonlinear dimension reduction for high dimensional approximation and inverse problems.

Alexandre PASCO

Centrale Nantes, Nantes Université, France
PhD thesis supervisor: Professor Anthony Nouy.

Introduction

Costly to evaluate $u : \mathcal{X} \to U$, parameter set $\mathcal{X} \subset \mathbb{R}^d$, Hilbert space U.

Offline: construct surrogate $\hat{u} : \mathcal{X} \to U$. **Online**: many evaluations of $\hat{u}$.

Model Order Reduction

- High-dim: $\dim(U) = n \gg 1$.
- "Low-dim" V_r approximating

$$\mathcal{M} := \{u(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\} \subset U.$$

- Linear case: reduced basis [Veroy et al., 2003],

$$\hat{u}(\mathbf{x}) = \sum_{i=1}^{r} a_i(\mathbf{x}) v^{(i)}, \quad v^{(i)} \in U.$$

Feature learning

- High-dim: $\dim(\mathcal{X}) = d \gg 1$.
- Approximate $u(\mathbf{x})$ by a function of $g(\mathbf{x}) \in \mathbb{R}^m$, $m \ll d$.
- Linear case: ridge function [Logan and Shepp, 1975],

$$\hat{u}(\mathbf{x}) = f(G^T \mathbf{x}),$$
$$G \in \mathbb{R}^{d \times m}, \quad f : \mathbb{R}^m \to U.$$

Introduction

Parameter-dependent high-dimensional linear equation,

$$\forall \mathbf{x} \in \mathcal{X}, \quad A(\mathbf{x})u(\mathbf{x}) = b(\mathbf{x}).$$

Inverse problem
Dictionary MOR

- $\mathbf{x}$ is unknown and "omitted".
- Linear measurements

$$\mathbf{z} = (\ell_1(u)), \cdots, \ell_m(u))).$$

- Dictionary $\mathcal{D}_K = \{v^{(i)}\}_{1 \leq i \leq K}$,

$$\hat{u}(\mathbf{x}) \cong \hat{u}(\mathbf{z}) = \sum_{i=1}^{r} a_{\alpha_i}(\mathbf{z}) v^{(\alpha_i)}.$$

Forward problem
Linear MOR

- $\mathbf{x}$ is known.
- Reduced space U_r is known.
- $A(\mathbf{x})$ is ill-conditioned.
- Preconditioner for better Galerkin projection and error estimation.

Inverse problem with MOR

- Linear MOR [Maday et al., 2015, Binev et al., 2017]: PBDW $\hat{u}_{U_r}(\mathbf{z})$,

$$\dim(U_r) = r \leq m, \quad \mathrm{dist}(U_r, \mathcal{M}) \leq \varepsilon_r, \quad \|u - \hat{u}_{U_r}(\mathbf{z})\|_U \leq \mu_r \varepsilon_r.$$

$+$ Online efficient. $\quad -$ Limited by linear Kolmogorov width.

- Piecewise linear MOR [Cohen et al., 2022]:

$$\mathcal{L}_r^N := \{U_r^{(i)} : 1 \leq i \leq N\}, \quad \mathrm{dist}(\bigcup_{1\leq i\leq N} U_r^{(i)}, \mathcal{M}) \leq \varepsilon_r.$$

Select $V^*(\mathbf{z})$ using $\mathcal{S}(\cdot, \mathcal{M})$ surrogate to $\mathrm{dist}(\cdot, \mathcal{M})$,

$$\min_{V \in \mathcal{L}_r^N} \mathcal{S}(\hat{u}_V(\mathbf{z}), \mathcal{M}), \quad \mathcal{S}(v, \mathcal{M}) := \min_{\mathbf{y} \in \mathcal{X}} \|A(\mathbf{y})v - b(\mathbf{y})\|_{U'}.$$

$+$ Near optimal selection for inf-sup stable A.
$+$ Nonlinear width [Temlyakov, 1998].
$+$ Parameter separable $\Rightarrow$ Online efficient.

Computing the selection criterion $\mathcal{S}$

- Assume parameter separability,

$$A(\mathbf{x}) = \sum_{q=1}^{d} x_q A_q \quad \text{and} \quad b(\mathbf{x}) = b_0.$$

- Computing $\mathcal{S}$ is a linear least-squares problem [Cohen et al., 2022],

$$\mathcal{S}(v, \mathcal{M}) = \min_{\mathbf{y} \in \mathcal{P}} \|G(v)\mathbf{y} - b_0\|_{U'}, \quad G(v) := (A_1 v, \cdots, A_{m_A} v).$$

- Offline: normal equations cost in total $\mathcal{O}(d^2 r^2 N n)$.
  + Online efficiency.
  − **Offline cost** may be prohibitive.
  − Sensitivity to **round-off errors**.

- **Random sketching** [Woodruff, 2014, Martinsson and Tropp, 2020] helps mitigate these problems [Balabanov and Nouy, 2019].

Subspace embeddings with random sketching

- $\Theta : U \to \mathbb{R}^k$ is a ϵ-**subspace embedding** for a subspace V if

$$\forall v \in V, \ \left| \|\Theta(v)\|_2^2 - \|v\|_U^2 \right| \leq \epsilon \|v\|^2$$

- For Θ with $k = \mathcal{O}(\epsilon^{-2}(\dim(V) + \log(\delta^{-1})))$ rows as independent **Gaussian** vectors with covariance depending on U,

$$\mathbb{P}\left[\Theta \text{ is subspace embedding for } V\right] \geq 1 - \delta.$$

  Θ is an **oblivious subspace embedding**.

- Similar for **structured** or **sparse** Θ with additional $\log$ terms.
  $\rightarrow$ Computing $\Theta(v)$ costs $n \log(n)$.

Sketched selection criterion

- With parameter separability and **structured** Θ,

$$\mathcal{S}^{\Theta}(v, \mathcal{M}) := \min_{\mathbf{y} \in \mathcal{X}} \|\Theta(A(\mathbf{y})v - b_0)\|_{U'} = \min_{\mathbf{y} \in \mathcal{X}} \|G^{\Theta}(v)\mathbf{y} - \Theta(b_0)\|_2,$$

$$G^{\Theta}(v) := (\Theta A_1 v, \cdots, \Theta A_d v) \in \mathbb{R}^{k \times d}.$$

- Offline: $\mathcal{S}^{\Theta}$ costs $\mathcal{O}(dr\log(n)Nn)$, while $\mathcal{S}$ costs $\mathcal{O}(d^2r^2Nn)$.
  $\rightarrow$ Lower offline cost.
  $\rightarrow$ More robust to round-off errors.

- Online: $\mathcal{S}^{\Theta}$ costs $\mathcal{O}(kd^2)$, while $\mathcal{S}$ costs $\mathcal{O}(d^3)$.

- Near optimal selection is preserved with high-probability and small k.

Proposition (Nouy and P. 2024)

With $k = \mathcal{O}(\epsilon^{-2}\ (d + \log(\delta^{-1})))$, for any $v \in U$,

$$\mathbb{P}\left[\sqrt{1-\epsilon}\,\mathcal{S}(v, \mathcal{P}) \leq \mathcal{S}^{\Theta}(v, \mathcal{P}) \leq \sqrt{1+\epsilon}\,\mathcal{S}(v, \mathcal{P})\right] \geq 1 - \delta.$$

Inverse problem with dictionary-based MOR

- Forward problem with dictionary-based MOR considered in [Kaulmann and Haasdonk, 2013, Balabanov and Nouy, 2021b].
- Given a dictionary $\mathcal{D}_K = \{v^{(1)}, \cdots, v^{(K)}\} \subset U$, consider

$$\mathcal{L}_r(\mathcal{D}_K) := \Big\{V \subset \operatorname{span}\{\mathcal{D}_K\} : \dim(V) \leq r\Big\}.$$

- Online adaptive library $\mathcal{L}_r(\mathbf{z}) \subset \mathcal{L}_r(\mathcal{D}_K)$ from greedy-type algorithm.
- **Offline cost**:

With random sketching $\mathcal{O}\big(mKn + dK\log(n)n\big)$. VS With normal equation $\mathcal{O}\big(mKn + d^2K^2n\big)$.

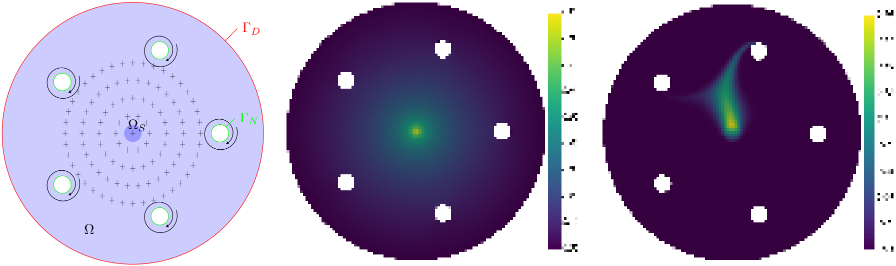
- **Online cost**: considering $k = \mathcal{O}(d)$,

With random sketching $\mathcal{O}\big(m^2K + md^2K + d^3K\big)$. VS With normal equation $\mathcal{O}\big(m^2K + m^2d^2K + d^3K\big)$.

Numerical experiment: parametrized advection diffusion

$n \sim 150\ 000$ and $\begin{cases} -0.01\Delta u + \mathcal{V}(\mathbf{x}) \cdot \nabla u = \dfrac{100}{\pi} \mathbb{1}_{\Omega_S} & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_D, \\ n \cdot \nabla u = 0 & \text{on } \Gamma_N, \end{cases}$

$$\mathcal{V}(\mathbf{x})(y) = \sum_{i=1}^{5} \frac{1}{\|y - y^{(i)}\|} \left(\mathbf{x}_i e_r(y^{(i)}) + \mathbf{x}_{i+5} e_\theta(y^{(i)}) \right)$$



Numerical experiment: parametrized advection diffusion

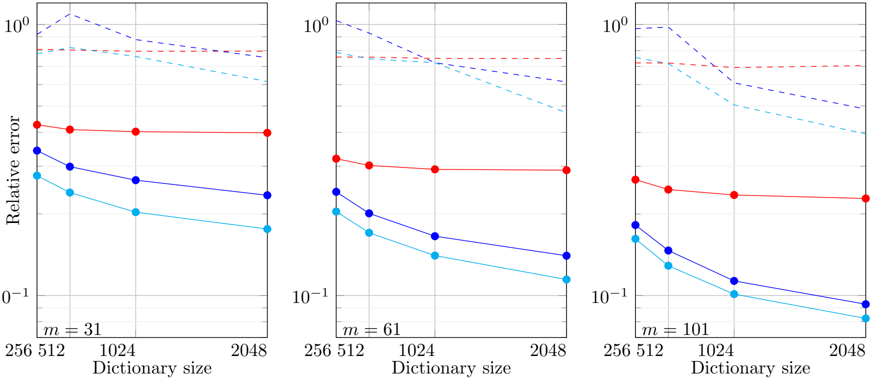


Figure: Relative H_0^1 errors on 500 test snapshots, growing dictionary sizes K, observations $m \in \{31, 61, 101\}$. Full line is the mean relative error. Dotted line is the maximal relative error. Blue is selection in $\mathcal{L}(w)$ with $\mathcal{S}^\Theta$. Red is $\min_{V \in \mathcal{L}^{POD}} \|u - \hat{u}_V\|_U$. Cyan is $\min_{V \in \mathcal{L}(w)} \|u - \hat{u}_V\|_U$.

Conclusion

Conclusion:

- Sketched selection criterion for piecewise-linear MOR.
- **Dictionary**-based MOR, tractable with **random sketching**.

Perspectives:

- Fix observations $(\ell_i)_{1 \leq i \leq m}$ and construct suited dictionary $\mathcal{D}_K$.
- Fix $\mathcal{D}_K$ and construct suited $(\ell_i)_{1 \leq i \leq m}$ (optimal information).
- Incorporate $\mathcal{S}^\Theta$ in the adaptive construction of $\mathcal{L}_r(\mathbf{z})$.

Introduction

- Forward problem: parameter $\mathbf{x} \in \mathcal{X}$ is known.

- MOR: U_r is given. Compute $u_r(\mathbf{x}) \in U_r$ via Galerkin projection,

$$\forall v \in U_r, \quad \langle v, A(\mathbf{x}) u_r(\mathbf{x}) \rangle = \langle v, b(\mathbf{x}) \rangle,$$

  and estimate error with $\widehat{\alpha}(A(\mathbf{x}))^{-1} \| A(\mathbf{x}) u_r(\mathbf{x}) - b(\mathbf{x}) \|_{U'}$.

- Problem: ill-conditioned operator $A(\mathbf{x})$.
  – $u_r(\mathbf{x})$ may be far from $\Pi_{U_r} u(\mathbf{x})$.
  – Residual-based error estimator may be not efficient.

- Goal: construct $P(\mathbf{x}) : U' \to U$ a linear approximation of $A(\mathbf{x})^{-1}$ [Zahm and Nouy, 2016, Balabanov and Nouy, 2021a],

$$P(\mathbf{x}) \in \mathrm{span}\{Y_1, \cdots, Y_p\}, \quad Y_i = A(\mathbf{x}_i)^{-1}.$$

Quality measure: general purpose in operator norm

- General purpose: discrepancy in operator norm $\|I - PA\|_{U,U}$.
- Bound on the inf-sup constants of PA,

$$1 - \|I - PA\|_{U,U} \leq \sigma_n(PA) \leq \sigma_1(PA) \leq 1 + \|I - PA\|_{U,U}.$$

- Preconditioned residual estimator $\|PAv - Pb\|_U$,

$$\frac{\|PAv - Pb\|_U}{1 + \|I - PA\|_{U,U}} \leq \|u - v\|_U \leq \frac{\|PAv - Pb\|_U}{1 - \|I - PA\|_{U,U}}.$$

- Problem: require $\|I - PA\|_{U,U} < 1$, but online computation and optimization of $P \mapsto \|I - PA\|_{U,U}$ is untractable.

Quality measure: general purpose in HS norm

- Alternative: Hilbert-Schmidt norm $\|I - PA\|_{HS(U,U)} \geq \|I - PA\|_{U,U}$.
- Minimization of the discrepancy is a least-squares problem,

$$\min_{P \in \text{span}\{Y_1, \cdots, Y_p\}} \|I - PA\|_{HS(U,U)}.$$

- Problem 1: fixed P, evaluating $\|I - PA\|_{HS(U,U)}$ is (very) costly.
  $\rightarrow$ [Zahm and Nouy, 2016] used random estimator.

- Problem 2: HS norm can be much larger than operator norm,

$$\frac{1}{\sqrt{\dim(U)}}\|\cdot\|_{HS(U,U)} \leq \|\cdot\|_{U,U} \leq \|\cdot\|_{HS(U,U)}.$$

$\rightarrow$ Seminorms tailored to MOR.

Quality measure: MOR purpose in operator seminorm

- MOR purpose: discrepancy in operator seminorms,

$$\|I - PA\|_{U,U_r} := \|\Pi_{U_r}(I - PA)\|_{U,U}.$$

- For u_r preconditioned Galerkin projection on U_r,

$$\|u - u_r\|_U \leq \frac{1}{1 - \|I - PA\|_{U,U_r}} \|u - \Pi_{U_r} u\|_U,$$

- Computable with $(v^{(i)})_{1 \leq i \leq r}$ o.n.b of U_r,

$$\|I - PA\|^2_{U,U_r} = \sigma_1\Big((\langle v^{(i)}, (I - PA)(I - PA)^T v^{(j)} \rangle_U)_{1 \leq i,j \leq r} \Big).$$

$+$ Online efficient using normal equation.
$-$ Offline is costly and sensitive to round-off errors.
$-$ Minimization over P is challenging.

Quality measure: MOR purpose in HS seminorm

- Assume given $U_m \supset U_r$, $r \leq m \ll n$, and for some $\tau \in (0, 1)$,

$$\|u - \Pi_{U_m} u\|_U \leq \tau \|u - u_r\|_U.$$

- A posteriori **error estimator**,

$$\frac{\|\Pi_{U_m} P(Au_r - b)\|_U}{1 + (1+\tau)\|I - PA\|_{U,U_m}} \leq \|u - u_r\|_U \leq \frac{\|\Pi_{U_m} P(Au_r - b)\|_U}{\sqrt{1-\tau^2} - (1+\tau)\|I - PA\|_{U,U_m}}.$$

- $U_r \subset U_m$ thus $\|\cdot\|_{U,U_r} \leq \|\cdot\|_{U,U_m}$.

Quality measure: MOR purpose in HS seminorm

- Alternative: Hilbert-Schmidt seminorm,

$$\|I - PA\|_{HS(U,U_m)} := \|\Pi_{U_m}(I - PA)\|_{HS(U,U)},$$

- HS seminorms are almost equivalent to operator seminorms,

$$\frac{1}{\sqrt{m}}\|\cdot\|_{HS(U,U_m)} \leq \|\cdot\|_{U,U_m} \leq \|\cdot\|_{HS(U,U_m)}.$$

- Computable with $(v^{(i)})_{1\leq i\leq m}$ o.n.b of U_m,

$$\|I - PA\|^2_{HS(U,U_m)} = \sum_{i=1}^{m} \|(I - PA)^T v^{(i)}\|^2_U,$$

+ Online efficient using normal equation.
− Offline is costly and sensitive to round-off errors.

Random sketching for HS operators

- Main challenge: Y_i are **inverses** of (sparse) matrices, only accessible via **matrix-vector products**.

- Random embedding: consider Ω, Σ and Γ "classical" random sketches, with sketching dimension k, then for an HS operator Y,

$$\Theta(Y) := \Gamma \text{vec}(\Omega Y \Sigma^T) \in \mathbb{R}^k.$$

$\rightarrow$ Computed with k matrix-vector products.
$\rightarrow$ Adaptable for seminorms.

- **Oblivious subspace embedding** with $k = \mathcal{O}(\epsilon^{-2}(d + \log(n/\delta)))$: for all d-dimensional vector subspace $\mathcal{Y}$ of HS operators,

$$\mathbb{P}\left[\forall Y \in \mathcal{Y}, \left| \|\Theta(Y)\|_2^2 - \|Y\|_{HS}^2 \right| \leq \epsilon \|Y\|_{HS}^2 \right] \geq 1 - \delta.$$

Practical aspects

- Assume parameter separability,

$$A(\mathbf{x}) = \sum_{q=1}^{d} x_q A_q.$$

- Construct $P(\mathbf{x})$ by solving the sketched least-squares problem,

$$\min_{P \in \text{span}\{Y_1, \cdots, Y_p\}} \|\Theta(I - PA(\mathbf{x}))\|_2 = \min_{a \in \mathbb{R}^p} \|\Theta(I) - W^{\Theta}(\mathbf{x})a\|_2,$$

  $W^{\Theta}(\mathbf{x}) \in \mathbb{R}^{k \times d}$ small matrix with **parameter separable** columns.

- If b is parameter separable, then preconditioned Galerkin system and preconditioned error estimator are also **parameter separable**.

- Offline cost for $HS(U, U_m)$: $\mathcal{O}((k \wedge m)pd \log(n)n)$.

Numerical example

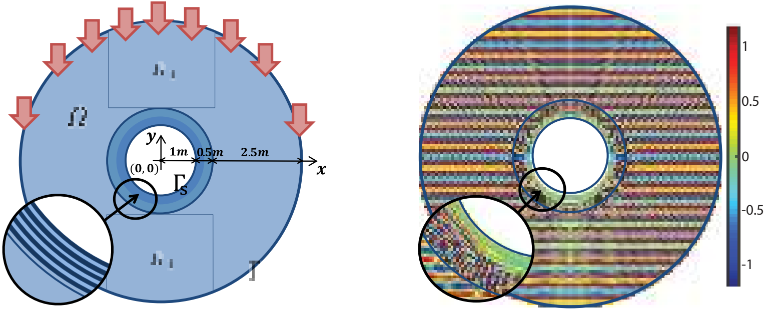


Figure: [Balabanov and Nouy, 2021b] Wave scattering in 2D with a perfect scatterer covered in an invisibility cloak composed of layers of homogeneous isotropic materials. Left: Geometry of the problem. Right: real part of random snapshot.

Numerical example

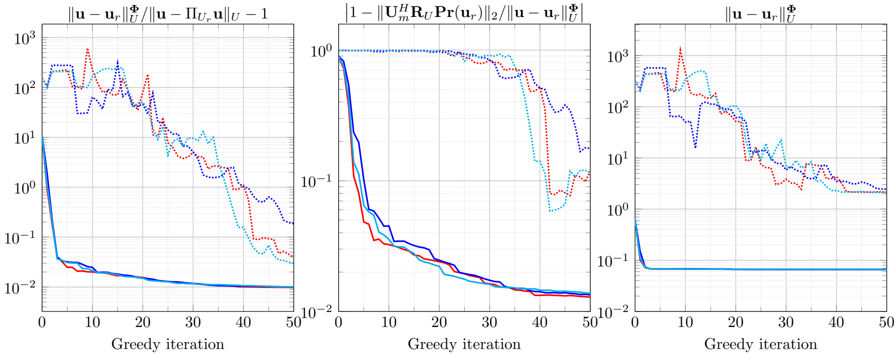


Figure: Quantiles on test sets (90% continuous, 100% dotted) for the preconditioned Galerkin projection, along the greedy construction of the preconditioner space. Left: quasi optimality. Middle: accuracy of error estimator. Right: absolute error to solution. Three sketched greedy criterions: red is $HS(U_m, U_m)$, blue is $HS(U, U_m)$, cyan is weighted sum.

Conclusion

Conclusion:

- Random sketching method for HS operators (generic approach).
- Construction of preconditioner for MOR purpose.
- Offline-online efficiency.

Perspectives for MOR:

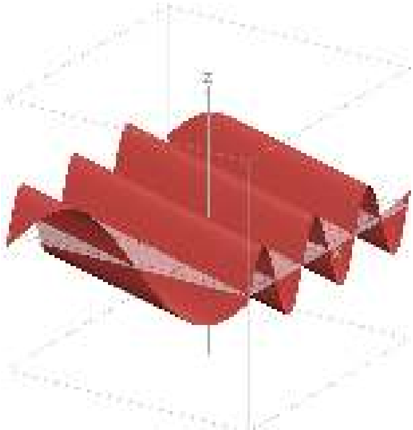
- Greedy algorithm constructing $(Y_i)_{1 \leq i \leq p}$ and U_r at the same time, as in [Zahm and Nouy, 2016].
- Nonlinear construction of $P(\mathbf{x})$ (e.g., dictionary-based).

Perspectives for random sketching:

- Sketching of operators for other settings (e.g., eigenvalue problems, domain decomposition).

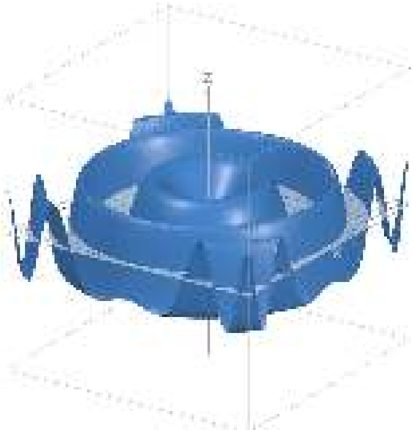
Introduction

$u_1(\mathbf{x}) = \sin(x_1 + 3x_2)$



Univariate in **linear feature**.

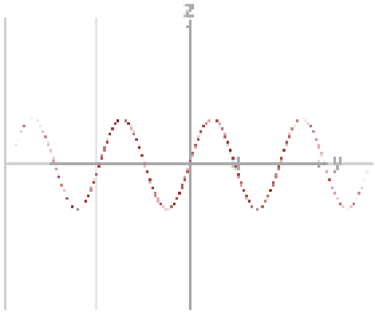
$u_2(\mathbf{x}) = \sin(x_1^2 + x_2^2)$



Univariate in **nonlinear feature**.

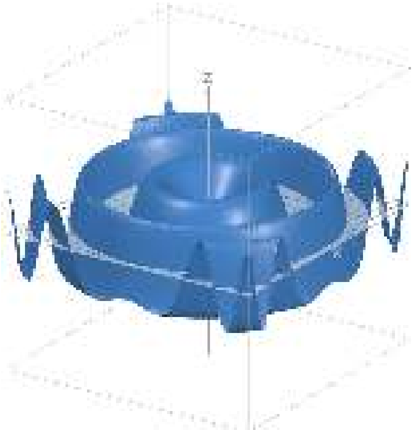
Introduction

$u_1(\mathbf{x}) = \sin(x_1 + 3x_2)$



Univariate in **linear feature**.

$u_2(\mathbf{x}) = \sin(x_1^2 + x_2^2)$



Univariate in **nonlinear feature**.

Introduction: gradient-based dimension reduction

- $\mathbf{X} \in \mathcal{X} \subset \mathbb{R}^d$ a random vector, $u : \mathcal{X} \to U = \mathbb{R}$, $d \gg 1$, $u \in \mathcal{C}^1$.
- Classical tools require sample size exponential in d.
  $\to$ Curse of dimensionality.
- Goal: build a **feature map** $g : \mathcal{X} \to \mathbb{R}^m$, $m \ll d$, so that u is well approximated by

  $$\hat{u} : \mathbf{x} \mapsto f(g(\mathbf{x}))$$

  for some low-dimensional **profile function** $f : \mathbb{R}^m \to \mathbb{R}$.
- Ideal measure of quality of g,

  $$\mathcal{E}(g) := \inf_{f:\mathbb{R}^m \to \mathbb{R}} \mathbb{E}\left[|u(\mathbf{X}) - f \circ g(\mathbf{X})|^2 \right].$$

- Given: few costly point evaluations $(\mathbf{x}^{(i)}, u(\mathbf{x}^{(i)}), \nabla u(\mathbf{x}^{(i)}))_{1 \leq i \leq n_s}$.

Examples of feature maps

Feature maps considered in gradient-based dimension reduction.

- [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]
  $g(\mathbf{x}) = G^T\Phi(\mathbf{x})$ from vector space with basis $\Phi : \mathbb{R}^d \to \mathbb{R}^K$ and $G \in \mathbb{R}^{K\times m}$ with $K \geq d$. Typically Φ polynomial.

- [Verdière et al., 2025, Zhang et al., 2019]
  Diffeomorphism-based $g(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_m(\mathbf{x}))^T$ with $\psi : \mathbb{R}^d \to \mathbb{R}^d$ a diffeomorphism.

Examples of feature maps

Feature maps considered in gradient-based dimension reduction.

- [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]
  $g(\mathbf{x}) = G^T\Phi(\mathbf{x})$ from vector space with basis $\Phi : \mathbb{R}^d \to \mathbb{R}^K$ and $G \in \mathbb{R}^{K\times m}$ with $K \geq d$. Typically Φ polynomial.

- [Verdière et al., 2025, Zhang et al., 2019]
  Diffeomorphism-based $g(\mathbf{x}) = (\psi_1(\mathbf{x}), \cdots, \psi_m(\mathbf{x}))^T$ with $\psi : \mathbb{R}^d \to \mathbb{R}^d$ a diffeomorphism.

Upper-bound using Poincaré inequalities

- For $g : \mathcal{X} \to \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}$ using the chain rule,

$$u = f \circ g \implies \nabla u(\mathbf{x}) = \nabla g(\mathbf{x}) \nabla f(g(\mathbf{x})),$$
$$\implies \nabla u(\mathbf{x}) \in \mathrm{span}\{\nabla g_1(\mathbf{x}), \cdots, \nabla g_m(\mathbf{x})\} \subset \mathbb{R}^d,$$

- We search for $g \in \mathcal{G}_m$ such that $\nabla g(\mathbf{x})$ is aligned with $\nabla u(\mathbf{x})$, for example by minimizing the objective function

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

Proposition ([Bigoni et al., 2022])

If ∇g has full matrix rank everywhere,

$$\mathcal{E}(g) \leq \Big(\sup_{h \in \mathcal{G}_m} \sup_{\mathbf{z} \in h(\mathrm{supp}\mathbf{X})} C_{\mathbf{X}|h(\mathbf{X})=\mathbf{z}} \Big) \mathcal{J}(g),$$

with $C_{\mathbf{X}|h(\mathbf{X})=\mathbf{z}}$ the Poincaré constant associated to the conditional measure $\mathbf{X}|h(\mathbf{X}) = \mathbf{z}$.

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: $g(\mathbf{x}) = G^T\mathbf{x}$ with $G^TG = I_m$, then $\Pi_{\nabla g(\mathbf{X})} = GG^T$, thus $G \mapsto \mathcal{J}(G^T)$ is **quadratic** and 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.
$\rightarrow$ Design quadratic surrogates.

Surrogate for $m = 1$

- With $m = 1$ the objective function writes

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

- Control on $\|\nabla g(\mathbf{X})\|_2^2$ yields control on $\mathcal{J}(g)$ with a **quadratic** surrogate,

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

- Uniform lower bound on $\|\nabla g(\mathbf{X})\|_2^2$ not available (e.g. polynomials).

Deviation inequalities for polynomials

Control $\|\nabla g(\mathbf{X})\|_2^2$ in terms of deviation inequalities, i.e. the decay of

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

Proposition (Direct consequence of [Fradelizi, 2009])

If $\mathbf{X}$ is uniformly distributed on a convex set and $h : \mathbb{R}^d \to \mathbb{R}$ is polynomial with total degree at most k, then for all $\beta > 0$,

$$\mathbb{P}\left[|h(\mathbf{X})| \leq \beta^{-1}\right] \lesssim \beta^{-1/k}.$$

- Generalized to s-concave measures (here $s = 1/d$).
- Generalized to h satisfying a Remez inequality.
- Constants behind $\lesssim$ involve moments of $|h(\mathbf{X})|$.

Analysis of the surrogate for $m = 1$

Assumptions

(1) $\mathbf{X}$ is uniformly distributed on a convex set.

(2) Every $g \in \mathcal{G}_1$ is a non-constant polynomial of total degree at most $\ell + 1$ such that $\mathbb{E}\left[\|\nabla g(\mathbf{X})\|_2^2\right] = 1$. $\rightarrow$ Feasible in practice.

Proposition (Nouy and P. 2025)

Under (1) *and* (2), *if* $\|\nabla u(\mathbf{X})\|_2^2 \leq 1$ *a.s., then for all* $g \in \mathcal{G}_m$,

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

with $\gamma_1 \leq 64 \min(3\ell, d)$ *and* $\gamma_2 \leq 2(8d)^{2\ell}$. *For* g^* *minimizer of* $\mathcal{L}_1$ *on* $\mathcal{G}_1$,

$$\mathcal{J}(g^*) \leq \gamma_3 \inf_{h \in \mathcal{G}_1} \mathcal{J}(h)^{\frac{1}{1+2\ell}}, \quad \gamma_3 \leq 1024 d \min(3\ell, d).$$

Results available for s-concave measures (here $s = 1/d$) and for functions satisfying a type of Remez inequality.

Minimizing the surrogate for $m = 1$

Assumption (2) is satisfied if we take $\Phi : \mathbb{R}^d \to \mathbb{R}^K$ polynomial such that $\nabla\Phi(\mathbf{X})$ has full rank a.s, and

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

Proposition (Nouy and P. 2025)

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

with $R := \mathbb{E}\left[\nabla\Phi(\mathbf{X})^T \nabla\Phi(\mathbf{X})\right] \in \mathbb{R}^{K \times K}$ *and*

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

Extension to $m > 1$

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

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

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

- Fix $g \in \mathcal{G}_m$, $h \mapsto \mathcal{L}_{m,j}((g_1, \cdots, g_{j-1}, h, g_{j+1}, \cdots, g_m))$ is **quadratic**.

Proposition (Nouy and P. 2025)

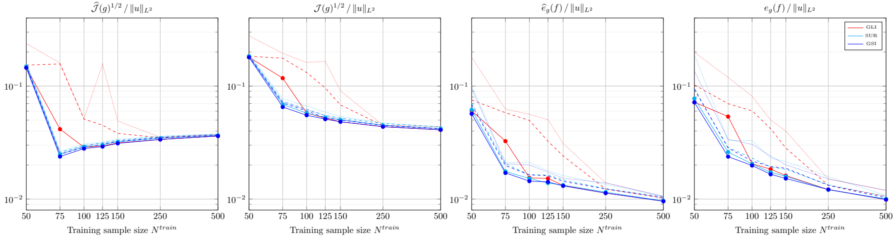
Under the previous assumptions, for all $g \in \mathcal{G}_m$,

$$\tilde{\gamma}_2^{-1} \mathcal{L}_{m,j}(g) \leq \mathcal{J}(g) \leq \tilde{\gamma}_1 \nu_{\mathcal{G}_m}^{-\frac{1}{1+2\ell m}} \mathcal{L}_{m,j}(g)^{\frac{1}{1+2\ell m}},$$

with $\tilde{\gamma}_1 \leq 2^9 m^{1/4\ell} d \min(d, 3\ell m)$, $\tilde{\gamma}_2 \leq 2^{7\ell} d^{2\ell}$ and

$$\nu_{\mathcal{G}_m} := \sup_{h \in \mathcal{G}_m} \mathbb{E}\left[\det \nabla h(\mathbf{X})^T \nabla h(\mathbf{X})\right].$$

Numerical experiment $m = 1$



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

50%, 90% and 100% quantiles over 20 realizations.

Left plots: train and test estimations of $\mathcal{J}(g)/\|u\|_{L^2}$.

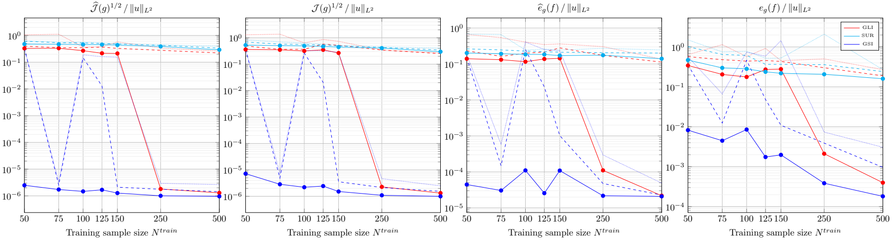
Right plots: train and test estimations of $\|u - f \circ g\|_{L^2}/\|u\|_{L^2}$.

Red: quasi-Newton minimization of $\mathcal{J}$.

Cyan: minimization of the surrogate.

Blue: surrogate as initialization of quasi-Newton for $\mathcal{J}$.

Numerical experiment $m = 2$



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

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 $\mathcal{J}$.

Introduction

Same setting as previous section, but with structured feature maps.

1- Collective setting, $Y \in \mathcal{Y}$ random variable, $u : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$,

$$\hat{u} : (\mathbf{x}, y) \mapsto f(g(\mathbf{x}), y).$$

2- Two variables setting, $\alpha \subset \{1, \cdots, d\}$, $u : \mathcal{X}_\alpha \times \mathcal{X}_{\alpha^c} \to \mathbb{R}$,

$$\hat{u} : \mathbf{x} \mapsto f(g^\alpha(\mathbf{x}_\alpha), g^{\alpha^c}(\mathbf{x}_{\alpha^c})).$$

3- Multiple variables setting, $S \subset \mathcal{P}(\{1, \cdots, d\})$, $u : \times_{\alpha \in S} \mathcal{X}_\alpha \to \mathbb{R}$,

$$\hat{u} : \mathbf{x} \mapsto f(g^{\alpha_1}(\mathbf{x}_{\alpha_1}), \cdots, g^{\alpha_{|S|}}(\mathbf{x}_{\alpha_{|S|}})).$$

Collective dimension reduction

- Goal: for $\mathbf{X} \in \mathcal{X} \subset \mathbb{R}^d$, $Y \in \mathcal{Y}$ independent of $\mathbf{X}$, approximate

$$u_Y := u(\cdot, Y) : \mathbb{R}^d \to \mathbb{R}, \quad u_Y \in \mathcal{C}^1(\mathcal{X}, \mathbb{R})$$

- Approximation format: use the same feature map for all Y,

$$\hat{u} : (\mathbf{x}, y) \mapsto f(g(\mathbf{x}), y).$$

- Ideal measure of quality of $g \in \mathcal{G}_m$,

$$\mathcal{E}_{\mathcal{X}}(g) := \inf_{f : \mathbb{R}^m \times \mathcal{Y} \to \mathbb{R}} \mathbb{E}\left[|u_Y(\mathbf{X}, Y) - f(g(\mathbf{X}), Y)|^2\right].$$

- Poincaré-based upper bound: apply [Bigoni et al., 2022] to u_Y,

$$\mathcal{E}_{\mathcal{X}}(g) \leq C_{\mathbf{X}|\mathcal{G}_m} \mathcal{J}_{\mathcal{X}}(g), \quad \mathcal{J}_{\mathcal{X}}(g) := \mathbb{E}\left[\|\Pi^{\perp}_{\nabla g(\mathbf{X})} \nabla u_Y(\mathbf{X})\|_2^2\right].$$

Truncation in the objective function

- Observe that

$$\mathcal{J}_{\mathcal{X}}(g) \geq \mathbb{E}_{\mathbf{X}} \left[\min_{V(\mathbf{X}) \in \mathbb{R}^{d \times m}} \mathbb{E}_{Y} \left[\| \Pi^{\perp}_{V(\mathbf{X})} \nabla u_{\mathbf{Y}}(\mathbf{X}) \|_2^2 \right] \right] := \varepsilon_m.$$

- For any $\mathbf{x} \in \mathcal{X}$, the solution $V_m(\mathbf{x})$ to the minimization problem is given by the dominant eigenspace of

$$M(\mathbf{X}) := \mathbb{E}_{Y} \left[\nabla u_{Y}(\mathbf{X}) \nabla u_{Y}(\mathbf{X})^T \right] \in \mathbb{R}^{d \times d}.$$

- Surrogate by truncating the part of $\nabla u_Y(\mathbf{X})$ orthogonal to $V_m(\mathbf{X})$,

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

Properties of the truncated objective function

- Minimizing $\mathcal{J}_{\mathcal{X},m}$ is almost the same as minimizing $\mathcal{J}_{\mathcal{X}}$, as

$$\frac{1}{2}(\mathcal{J}_{\mathcal{X},m}(g) + \varepsilon_m) \leq \mathcal{J}_{\mathcal{X}}(g) \leq \mathcal{J}_{\mathcal{X},m}(g) + \varepsilon_m.$$

- $\mathcal{J}_{\mathcal{X},m}$ is better suited for our construction of quadratic surrogates,

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

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

- Need concentration inequalities on $\sigma_i(\nabla g(\mathbf{X}))^2$.

Surrogate for collective dimension reduction

- Define surrogate in collective setting,

$$\mathcal{L}_{\mathcal{X},m}(g) := \mathbb{E}\left[\sigma_1(M(\mathbf{X}))\|\Pi^{\perp}_{V_m(\mathbf{X})}\nabla g(\mathbf{X})\|_F^2\right].$$

Proposition (Nouy and P. 2025)

Under (1) *and* (2), *if* $\sigma_1(M(\mathbf{X})) \leq 1$ *a.s., then for all* $g \in \mathcal{G}_m$,

$$\mathcal{J}_{\mathcal{X},m}(g) \leq \gamma_{\mathcal{G}_m}\mathcal{L}_{\mathcal{X},m}(g)^{\frac{1}{1+2\ell m}}.$$

- The surrogate is quadratic, $\mathcal{L}_{\mathcal{X},m}(G^T\Phi) = \mathrm{Tr}\left(G^T H_{\mathcal{X},m} G\right)$, with

$$H_{\mathcal{X},m} = \mathbb{E}\left[\sigma_1(M(\mathbf{X}))\nabla\Phi(\mathbf{X})(I_d - V_m(\mathbf{X})V_m(\mathbf{X})^T)\nabla\Phi(\mathbf{X})^T\right].$$

- Problem: estimating $M(\mathbf{x}) = \mathbb{E}_{Y}\left[\nabla u_{Y}(\mathbf{x})\nabla u_{Y}(\mathbf{x})^T\right]$ and dominant eigenvectors $V_m(\mathbf{x})$. Requires specific sampling (tensorized).

Two variables approach

- Goal: for $\alpha \subset \{1, \cdots, d\}$, split $\mathbf{X} = (\mathbf{X}_\alpha, \mathbf{X}_{\alpha^c})$, assume $\mathbf{X}_\alpha \perp\!\!\!\perp \mathbf{X}_{\alpha^c}$, approximate $u : \mathbb{R}^d \to \mathbb{R} \in \mathcal{C}^1$.

- Approximation format: separated features in $\mathbf{X}_\alpha$ and $\mathbf{X}_{\alpha^c}$,

$$\hat{u} : \mathbf{x} \mapsto f(g^\alpha(\mathbf{x}_\alpha), g^{\alpha^c}(\mathbf{x}_{\alpha^c})).$$

- Ideal measure of quality of $g \in \mathcal{G}_m$ writes

$$\mathcal{E}(g) = \inf_{f:\mathbb{R}^{m_\alpha} \times \mathbb{R}^{m_{\alpha^c}} \to \mathbb{R}} \mathbb{E}\left[|u_Y(\mathbf{X}) - f(g^\alpha(\mathbf{X}_\alpha), g^{\alpha^c}(\mathbf{X}_{\alpha^c}))|^2 \right].$$

- Link to the collective setting to use our surrogates ?

Two variables approach

- Recall that $g : \mathbf{x} \mapsto (g^{\alpha}(\mathbf{x}_{\alpha}), g^{\alpha^c}(\mathbf{x}_{\alpha^c}))$.

- Definition of the Poincaré inequality based objective function yields

$$\mathcal{J}(g) = \mathcal{J}_{\mathcal{X}_{\alpha}}(g^{\alpha}) + \mathcal{J}_{\mathcal{X}_{\alpha^c}}(g^{\alpha^c}).$$

$\Rightarrow$ Exactly the same as two collective settings.

- Inspiring from analysis of HOSVD for Tucker tensor format,

$$\mathcal{E}(g) \leq \mathcal{E}_{\mathcal{X}_{\alpha}}(g^{\alpha}) + \mathcal{E}_{\mathcal{X}_{\alpha^c}}(g^{\alpha^c}) \leq 2\mathcal{E}(g).$$

$\Rightarrow$ Almost the same as two collective settings.

Multiple variables approach

- Consider $g : \mathbf{x} \mapsto (g^{\alpha_1}(\mathbf{x}_{\alpha_1}), \cdots, g^{\alpha_{|S|}}(\mathbf{x}_{\alpha_{|S|}}))$, $S \subset \mathcal{P}(\{1, \cdots, d\})$.

- Definition of the Poincaré inequality based objective function yields

$$\mathcal{J}(g) = \sum_{\alpha \in S} \mathcal{J}_{\mathcal{X}_\alpha}(g^\alpha).$$

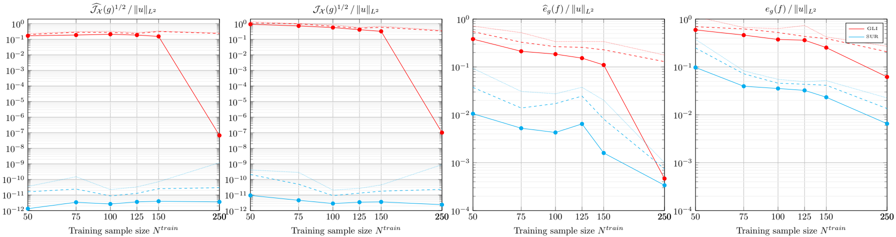
$\Rightarrow$ Exactly the same as $|S|$ collective settings.

- Approach from HOSVD of Tucker tensor format and independence of $(\mathbf{X}_\alpha)_{\alpha \in S}$ yields

$$\mathcal{E}(g) \leq \sum_{\alpha \in S} \mathcal{E}_{\mathcal{X}_\alpha}(g^\alpha) \leq |S| \mathcal{E}(g).$$

$\Rightarrow$ Almost the same as $|S|$ collective settings.

Numerical experiment: collective setting



$$u(\mathbf{x}, y) := \sum_{k=1}^{p} (\mathbf{x}^T Q_k \mathbf{x})^2 \sin(\frac{\pi k}{2p} y),$$

$$\mathbf{X} \sim \mathcal{U}(]-1,1[^8), \quad Y \sim \mathcal{U}(]-1,1[), \quad p = m = 3.$$

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.

Conclusion

Conclusion:

- **Quadratic surrogate** to the non-convex objective function arising from **gradient-based dimension reduction**.
- Quasi-optimality results for our surrogates, especially for $m = 1$.
- Extension to the **collective dimension reduction** setting.
- Correspondence between separated features and collective setting.

Perspectives:

- Other classes of feature maps (e.g. diffeomorphisms, low-rank).
- Extensive numerical tests.
- Applications to compression of fast-to-evaluate functions u.
- Extension to Bayesian inverse problems.

Thank you for your attention.

References I

Balabanov, O. and Nouy, A. (2019).
Randomized linear algebra for model reduction. Part I: Galerkin methods and error estimation.
Adv Comput Math, 45(5-6):2969–3019.

Balabanov, O. and Nouy, A. (2021a).
Preconditioners for model order reduction by interpolation and random sketching of operators.
arXiv:2104.12177 [cs, math].

Balabanov, O. and Nouy, A. (2021b).
Randomized linear algebra for model reduction—part II: Minimal residual methods and dictionary-based approximation.
Adv Comput Math, 47(2):26.

References II

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.

Binev, P., Cohen, A., Dahmen, W., DeVore, R., Petrova, G., and Wojtaszczyk, P. (2017).
Data Assimilation in Reduced Modeling.
SIAM/ASA J. Uncertainty Quantification, 5(1):1–29.

Cohen, A., Dahmen, W., Mula, O., and Nichols, J. (2022).
Nonlinear Reduced Models for State and Parameter Estimation.
SIAM/ASA J. Uncertainty Quantification, 10(1):227–267.

References III

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).

Kaulmann, S. and Haasdonk, B. (2013).
Online greedy reduced basis construction using dictionaries.
In *VI International Conference on Adaptive Modeling and Simulation ADMOS 2013*, page 12.

References IV

Logan, B. F. and Shepp, L. A. (1975).
Optimal reconstruction of a function from its projections.
Duke Math. J., 42(4).

Maday, Y., Patera, A. T., Penn, J. D., and Yano, M. (2015).
A parameterized-background data-weak approach to variational data assimilation: Formulation, analysis, and application to acoustics.
Int. J. Numer. Meth. Engng, 102(5):933–965.

Martinsson, P.-G. and Tropp, J. A. (2020).
Randomized numerical linear algebra: Foundations and algorithms.
Acta Numerica, 29:403–572.

References V

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.

Temlyakov, V. N. (1998).
Nonlinear Kolmogorov widths.
Math Notes, 63(6):785–795.

Verdière, R., Prieur, C., and Zahm, O. (2025).
Diffeomorphism-based feature learning using Poincaré inequalities on augmented input space.
Journal of Machine Learning Research, 26(139):1–31.

References VI

Veroy, K., Prud'homme, C., Rovas, D., and Patera, A. (2003).
A Posteriori Error Bounds for Reduced-Basis Approximation of Parametrized Noncoercive and Nonlinear Elliptic Partial Differential Equations.
In *16th AIAA Computational Fluid Dynamics Conference*, Orlando, Florida. American Institute of Aeronautics and Astronautics.

Woodruff, D. P. (2014).
Sketching as a Tool for Numerical Linear Algebra.
FNT in Theoretical Computer Science, 10(1-2):1–157.

Zahm, O. and Nouy, A. (2016).
Interpolation of Inverse Operators for Preconditioning Parameter-Dependent Equations.
SIAM J. Sci. Comput., 38(2):A1044–A1074.

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.