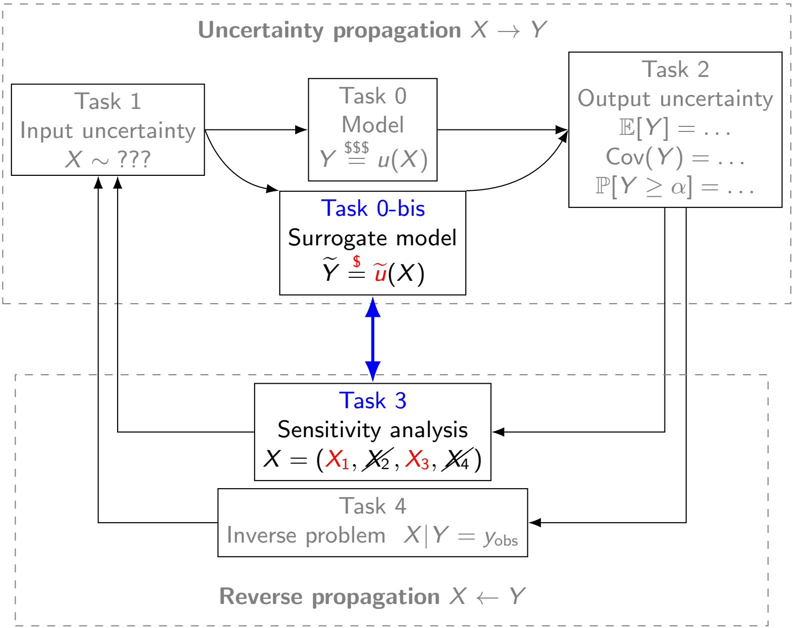
Poincaré inequalities for dimension reduction and efficient sampling

Part II: Gradient-based dimension reduction

Olivier Zahm
olivier.zahm@inria.fr

INRIA – Université Grenoble-Alpes
https://team.inria.fr/airsea/en/

ETICS
October 2025


Uncertainty propagation $X \to Y$
Task 1
Input uncertainty
$X \sim ???$
Task 0
Model
$Y \overset{\$\$\$}{=} u(X)$
Task 2
Output uncertainty
$\mathbb{E}[Y] = \dots$
$\mathrm{Cov}(Y) = \dots$
$\mathbb{P}[Y \geq \alpha] = \dots$
Task 0-bis
Surrogate model
$\widetilde{Y} \overset{\$}{=} \widetilde{u}(X)$
Task 3
Sensitivity analysis
$X = (X_1, \cancel{X_2}, X_3, \cancel{X_4})$
Task 4
Inverse problem $X|Y = y_{\text{obs}}$
Reverse propagation $X \leftarrow Y$

Remember: **regularity is not enough!**

High-dimensional function approximation requires leveraging any **underlying low-dimensional structure** present in u, such as:

- **Sparse structure**

$$u(x) \approx \sum_{\alpha \in \Lambda_N} u_\alpha \Phi_\alpha(x), \qquad \begin{cases} \Phi_\alpha(x) : \text{ given basis} \\ \#\Lambda_N = N \end{cases}$$

- **Low-rank structure**

$$u(x) \approx \sum_{i=1}^{m} u_1^i(x_1) \dots u_d^i(x_d)$$

- **Compositional structure**

$$\boxed{u(x) \approx f \circ g(x), \qquad \begin{cases} g : \mathbb{R}^d \to \mathbb{R}^m \\ f : \mathbb{R}^m \to \mathbb{R} \\ m : \text{ latent dimension} \end{cases}}$$

- **...**

Low-effective dimension: latent dimension $m \ll d$

Given $u : \mathbb{R}^d \to \mathbb{R}$, find a **feature map** $g : \mathbb{R}^d \to \mathbb{R}^m$ and a **profile function** $f : \mathbb{R}^m \to \mathbb{R}$ with **latent dimension** $m \ll d$ such that

$$\mathbb{E}[(u(X) - f \circ g(X))^2] \leq \varepsilon$$

(regression problem in L^2)

Low-effective dimension: latent dimension $m \ll d$

Given $u : \mathbb{R}^d \to \mathbb{R}$, find a **feature map** $g : \mathbb{R}^d \to \mathbb{R}^m$ and a **profile function** $f : \mathbb{R}^m \to \mathbb{R}$ with **latent dimension** $m \ll d$ such that

$$\mathbb{E}[(u(X) - f \circ g(X))^2] \leq \varepsilon$$

(regression problem in L^2)

A trivial solution: $\{g = u \,;\, f = id \,;\, m = 1\}$ yields $\boxed{u = f \circ g}$. We have to restrict g in a **tractable** class

$$g \in \mathcal{G}$$

Low-effective dimension: latent dimension $m \ll d$

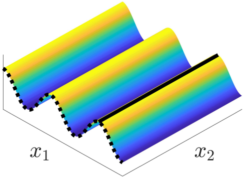
Given $u : \mathbb{R}^d \to \mathbb{R}$, find a **feature map** $g : \mathbb{R}^d \to \mathbb{R}^m$ and a **profile function** $f : \mathbb{R}^m \to \mathbb{R}$ with **latent dimension** $m \ll d$ such that

$$\mathbb{E}[(u(X) - f \circ g(X))^2] \leq \varepsilon$$

(regression problem in L^2)

A trivial solution: $\{g = u \; ; \; f = id \; ; \; m = 1\}$ yields $\boxed{u = f \circ g}$. We have to restrict g in a **tractable** class

$$g \in \mathcal{G}$$



$$\sin\big(\underbrace{x_1}\big)$$

$\mathcal{G}$ = coordinate selection
(sensitivity analysis)

Low-effective dimension: latent dimension $m \ll d$

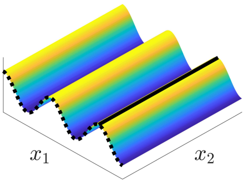
Given $u : \mathbb{R}^d \to \mathbb{R}$, find a **feature map** $g : \mathbb{R}^d \to \mathbb{R}^m$ and a **profile function** $f : \mathbb{R}^m \to \mathbb{R}$ with **latent dimension** $m \ll d$ such that

$$\mathbb{E}[(u(X) - f \circ g(X))^2] \leq \varepsilon$$

(regression problem in L^2)

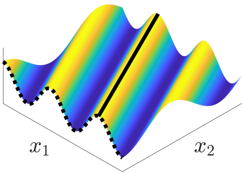
A trivial solution: $\{g = u \; ; \; f = id \; ; \; m = 1\}$ yields $\boxed{u = f \circ g}$. We have to restrict g in a **tractable** class

$$g \in \mathcal{G}$$



$$\sin\big(\underbrace{x_1}\big)$$

$\mathcal{G}$ = coordinate selection (sensitivity analysis)



$$\sin\big(\underbrace{x_1 + x_2}\big)$$

$\mathcal{G}$ = linear projection (active subspace)

Low-effective dimension: latent dimension $m \ll d$

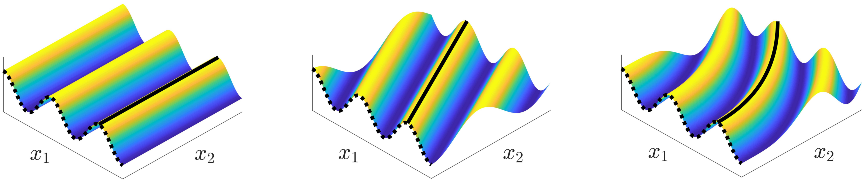
Given $u : \mathbb{R}^d \to \mathbb{R}$, find a **feature map** $g : \mathbb{R}^d \to \mathbb{R}^m$ and a **profile function** $f : \mathbb{R}^m \to \mathbb{R}$ with **latent dimension** $m \ll d$ such that

$$\mathbb{E}[(u(X) - f \circ g(X))^2] \le \varepsilon$$

(regression problem in L^2)

A trivial solution: $\{g = u \; ; \; f = id \; ; \; m = 1\}$ yields $\boxed{u = f \circ g}$. We have to restrict g in a **tractable** class

$g \in \mathcal{G}$



$\sin\big(\underbrace{x_1}\big)$	$\sin\big(\underbrace{x_1 + x_2}\big)$	$\sin\big(\underbrace{x_1 + x_2^2}\big)$
$\mathcal{G}$ = coordinate selection (sensitivity analysis)	$\mathcal{G}$ = linear projection (active subspace)	$\mathcal{G}$ = nonlinear extraction (active manifold)

Gradient-based dimension reduction

$$u(x)' = 0 \qquad \Leftrightarrow \qquad u = constant$$

$$u(x)' = 0 \qquad \Leftrightarrow \qquad u = constant$$

$$\partial_2 u(x_1, x_2) = 0 \qquad \Leftrightarrow \qquad u(x_1, x_2) = u(x_1)$$

$$u(x)' = 0 \qquad \Leftrightarrow \qquad u = constant$$

$$\partial_2 u(x_1, x_2) = 0 \qquad \Leftrightarrow \qquad u(x_1, x_2) = u(x_1)$$

$$\mathbb{E}[(\partial_2 u(X_1, X_2))^2] \approx 0 \qquad \overset{???}{\Leftrightarrow} \qquad u(x_1, x_2) \approx f(x_1)$$

$$u(x)' = 0 \qquad \Leftrightarrow \qquad u = constant$$

$$\partial_2 u(x_1, x_2) = 0 \qquad \Leftrightarrow \qquad u(x_1, x_2) = u(x_1)$$

$$\mathbb{E}[(\partial_2 u(X_1, X_2))^2] \approx 0 \qquad \overset{???}{\Leftrightarrow} \qquad u(x_1, x_2) \approx f(x_1)$$

Poincaré inequality (PI)

A probability measure π_X on $\mathbb{R}^d$ satisfies the Poincaré inequality with constant $\mathbb{C}(\pi_X) < \infty$ if

$$\mathrm{Var}(u(X)) \leq \mathbb{C}(\pi_X)\ \mathbb{E}[\|\nabla u(X)\|^2]$$

holds for any smooth function $u : \mathbb{R}^d \to \mathbb{R}$.

$$u(x)' = 0 \qquad \Leftrightarrow \qquad u = constant$$

$$\partial_2 u(x_1, x_2) = 0 \qquad \Leftrightarrow \qquad u(x_1, x_2) = u(x_1)$$

$$\mathbb{E}[(\partial_2 u(X_1, X_2))^2] \approx 0 \qquad \overset{???}{\Leftrightarrow} \qquad u(x_1, x_2) \approx f(x_1)$$

Poincaré inequality (PI)

A probability measure π_X on $\mathbb{R}^d$ satisfies the Poincaré inequality with constant $\mathbb{C}(\pi_X) < \infty$ if

$$\mathrm{Var}(u(X)) \leq \mathbb{C}(\pi_X)\ \mathbb{E}[\|\nabla u(X)\|^2]$$

holds for any smooth function $u : \mathbb{R}^d \to \mathbb{R}$.

If π_X satisfies

$$\overline{\mathbb{C}}(\pi_X) := \sup_{x_1} \mathbb{C}(\pi_{X_2|X_1=x_1}) < \infty$$

then there exists $f(X_1)$ (the conditional expectation $f(X_1) = \mathbb{E}[u(X)|X_1]$) such that

$$\boxed{\mathbb{E}[(u(X) - f(X_1))^2] \leq \overline{\mathbb{C}}(\pi_X)\, \mathbb{E}[(\partial_2 u(X))^2]}$$

Proof

$$
\begin{aligned}
\min_f \mathbb{E}[(u(X_1, X_2) - f(X_1))^2] &= \mathbb{E}\left[(u(X_1, X_2) - \mathbb{E}[u(X_1, X_2)|X_1])^2\right] \\
\text{(tower property)} \qquad &= \mathbb{E}\left[\mathbb{E}[(u(X_1, X_2) - \mathbb{E}[u(X_1, X_2)|X_1])^2|X_1]\right] \\
\text{(def. conditional variance)} \qquad &= \mathbb{E}\Big[\mathrm{Var}(u(X_1, X_2)|X_1)\Big] \\
\text{(Poincaré on } X_2|X_1) \qquad &\leq \mathbb{E}\left[\mathbb{C}(\pi_{X_2|X_1})\mathbb{E}[(\partial_2 u(X_1, X_2))^2|X_1]\right] \\
\text{(definition of } \overline{\mathbb{C}}(\pi_X)) \qquad &\leq \overline{\mathbb{C}}(\pi_X)\mathbb{E}\left[\mathbb{E}[(\partial_2 u(X_1, X_2))^2|X_1]\right] \\
\text{(tower property again)} \qquad &= \overline{\mathbb{C}}(\pi_X)\mathbb{E}[(\partial_2 u(X_1, X_2))^2]
\end{aligned}
$$

Generalization

Given a **unitary matrix** $U = [U_m, U_\perp] \in \mathbb{R}^{d \times d}$, we decompose X as

$$X = U_m X_m + U_\perp X_\perp \qquad \begin{cases} X_m = U_m^\top X & \in \mathbb{R}^m \\ X_\perp = U_\perp^\top X & \in \mathbb{R}^{d-m} \end{cases}$$

Subspace Poincaré inequality

A probability measure π_X on $\mathbb{R}^d$ satisfies the subspace Poincaré inequality with constant $\overline{\mathbb{C}}(\pi_X) < \infty$ if

$$\overline{\mathbb{C}}(\pi_X) = \sup_{\substack{U_m, U_\perp \text{ such that} \\ [U_m, U_\perp] \text{ is unitary}}} \sup_{x_m \in \mathbb{R}^m} \mathbb{C}(\pi_{X_\perp | X_m = x_m}) < \infty$$

Generalization

Given a **unitary matrix** $U = [U_m, U_\perp] \in \mathbb{R}^{d \times d}$, we decompose X as

$$X = U_m X_m + U_\perp X_\perp \qquad \begin{cases} X_m = U_m^\top X & \in \mathbb{R}^m \\ X_\perp = U_\perp^\top X & \in \mathbb{R}^{d-m} \end{cases}$$

Subspace Poincaré inequality

A probability measure π_X on $\mathbb{R}^d$ satisfies the subspace Poincaré inequality with constant $\overline{\mathbb{C}}(\pi_X) < \infty$ if

$$\overline{\mathbb{C}}(\pi_X) = \sup_{\substack{U_m, U_\perp \text{ such that} \\ [U_m, U_\perp] \text{ is unitary}}} \sup_{x_m \in \mathbb{R}^m} \mathbb{C}(\pi_{X_\perp | X_m = x_m}) < \infty$$

For any matrix U_m, there exists f such that

$$\boxed{\begin{aligned} \mathbb{E}[(u(X) - f(U_m^\top X))^2] &\leq \overline{\mathbb{C}}(\pi_X) \mathbb{E}[\| U_\perp^\top \nabla u(X) \|^2] \\ &= \overline{\mathbb{C}}(\pi_X) \Big(\mathbb{E}[\| \nabla u(X) \|^2] - \mathbb{E}[\| U_m^\top \nabla u(X) \|^2] \Big) \end{aligned}}$$

The RHS is quadratic in U_m, hence simple to minimize (compared to the LHS).

Active Subspace (AS)[1] = PCA on $\nabla u(X)$

$$\min_f \mathbb{E}[(u(X) - f(U_m^\top X))^2] \leq \overline{\mathbb{C}}(\pi_X)\Big(\mathbb{E}[\|\nabla u(X)\|^2] - \mathbb{E}[\|U_m^\top \nabla u(X)\|^2]\Big)$$
$$= \overline{\mathbb{C}}(\pi_X)\Big(\mathrm{trace}(H) - \mathrm{trace}(U_m^\top H U_m)\Big)$$

where H is the **active subspace** matrix

$$\boxed{H = \mathbb{E}[\nabla u(X) \nabla u(X)^\top]}$$

[1] [Constantine et.al. 2014]

Active Subspace (AS)[1] = PCA on $\nabla u(X)$

$$\min_f \mathbb{E}[(u(X) - f(U_m^\top X))^2] \leq \overline{\mathbb{C}}(\pi_X)\Big(\mathbb{E}[\|\nabla u(X)\|^2] - \mathbb{E}[\|U_m^\top \nabla u(X)\|^2]\Big)$$
$$= \overline{\mathbb{C}}(\pi_X)\Big(\mathrm{trace}(H) - \mathrm{trace}(U_m^\top H U_m)\Big)$$

where H is the **active subspace** matrix

$$\boxed{H = \mathbb{E}[\nabla u(X)\nabla u(X)^\top]}$$

Minimizing the RHS corresponds to **compute a PCA on $\nabla u(X)$**

1. Compute H *e.g.* via Monte Carlo: $H^N = \frac{1}{N}\sum_{i=1}^N \nabla u(X^{(i)})\nabla u(X^{(i)})^\top$
2. Solve the eigenvalue problem $\boxed{Hv_i = \lambda_i v_i}$ with $\lambda_1 \geq \lambda_2 \geq ... \geq 0$ and assemble
$$U_m = [v_1, \dots, v_m]$$
$$U_\perp = [v_{m+1}, \dots, v_d]$$
3. Finally we obtain
$$\min_f \mathbb{E}[(u(X) - f(U_m^\top X))^2] \leq \overline{\mathbb{C}}(\pi_X) \sum_{i=m+1}^{d} \lambda_i$$
   - Choose m such that the RHS is bellow a prescribed tolerance,
   - Compute f (polynomial, tensors, KRR, NN,...)

[1] [Constantine et.al. 2014]

Link with **Sobol' indices**

Instead of minimizing over U_m, take $U_m = [e_i]_{i\in\tau}$ such that $\boxed{U_m^\top X = X_\tau}$ for any set of indices $\tau \subset \{1, \dots, d\}$

$$\min_f \mathbb{E}[(u(X) - f(X_\tau))^2] \leq \overline{\mathbb{C}}(\pi_X)\Big(\mathrm{trace}(H) - \mathrm{trace}(U_m^\top H U_m)\Big)$$
$$= \overline{\mathbb{C}}(\pi_X) \sum_{i\in -\tau} H_{ii}$$

Link with **Sobol' indices**

Instead of minimizing over U_m, take $U_m = [e_i]_{i\in\tau}$ such that $\boxed{U_m^\top X = X_\tau}$ for any set of indices $\tau \subset \{1, \ldots, d\}$

$$\min_f \mathbb{E}[(u(X) - f(X_\tau))^2] \leq \overline{\mathbb{C}}(\pi_X)\Big(\text{trace}(H) - \text{trace}(U_m^\top H U_m)\Big) = \overline{\mathbb{C}}(\pi_X) \sum_{i\in-\tau} H_{ii}$$

Then we get

$$\boxed{T_\tau \leq \frac{\overline{\mathbb{C}}(\pi_X)}{\text{Var}(Y)} \sum_{i\in\tau} H_{ii}} \quad \text{and} \quad \boxed{S_\tau \geq 1 - \frac{\overline{\mathbb{C}}(\pi_X)}{\text{Var}(Y)} \sum_{i\in-\tau} H_{ii}}$$

where

$$H_{ii} = \mathbb{E}\left[(\partial_i u(X))^2\right]$$

are called the **Derivative based Global Sensitivity Measure** (DGSM).

Link with **Sobol' indices**

Instead of minimizing over U_m, take $U_m = [e_i]_{i\in\tau}$ such that $\boxed{U_m^\top X = X_\tau}$ for any set of indices $\tau \subset \{1,\dots,d\}$

$$\min_f \mathbb{E}[(u(X) - f(X_\tau))^2] \leq \overline{\mathbb{C}}(\pi_X)\Big(\mathrm{trace}(H) - \mathrm{trace}(U_m^\top H U_m)\Big)$$
$$= \overline{\mathbb{C}}(\pi_X) \sum_{i\in-\tau} H_{ii}$$

Then we get

$$\boxed{T_\tau \leq \frac{\overline{\mathbb{C}}(\pi_X)}{\mathrm{Var}(Y)} \sum_{i\in\tau} H_{ii}} \quad \text{and} \quad \boxed{S_\tau \geq 1 - \frac{\overline{\mathbb{C}}(\pi_X)}{\mathrm{Var}(Y)} \sum_{i\in-\tau} H_{ii}}$$

where

$$H_{ii} = \mathbb{E}\left[(\partial_i u(X))^2\right]$$

are called the **Derivative based Global Sensitivity Measure** (DGSM).

Remark: for independent inputs $X_i \perp X_j$, we even get [Zahm et.al, 2020]

$$\boxed{T_\tau \leq \sum_{i\in\tau} \frac{\mathbb{C}(\pi_{X_i})}{\mathrm{Var}(Y)} H_{ii}} \quad \text{and} \quad \boxed{S_\tau \geq 1 - \sum_{i\in-\tau} \frac{\mathbb{C}(\pi_{X_i})}{\mathrm{Var}(Y)} H_{ii}}$$

How to **control** $\overline{\mathbb{C}}(\pi_X)$? [Zahm et al, 2022]

Lower-bound (easy): by testing $\mathrm{Var}(f(X)) \leq \mathbb{C}(\pi_X)\mathbb{E}[\|\nabla f(x)\|^2]$ with affine functions, we get

$$\sup_{f:\text{affine}} \frac{\mathrm{Var}(f(X))}{\mathbb{E}[\|\nabla f(X)\|^2]} = \lambda_{\max}(\mathrm{Cov}(X)) \leq \mathbb{C}(\pi_X) \overset{\text{(definition of subspace PI)}}{\leq} \overline{\mathbb{C}}(\pi_X)$$

How to **control** $\overline{\mathbb{C}}(\pi_X)$? [Zahm et al, 2022]

Lower-bound (easy): by testing $\mathrm{Var}(f(X)) \leq \mathbb{C}(\pi_X)\mathbb{E}[\|\nabla f(x)\|^2]$ with affine functions, we get

$$\sup_{f:\text{affine}} \frac{\mathrm{Var}(f(X))}{\mathbb{E}[\|\nabla f(X)\|^2]} = \lambda_{\max}(\mathrm{Cov}(X)) \leq \mathbb{C}(\pi_X) \overset{\substack{\text{(definition of}\\ \text{subspace PI)}}}{\leq} \overline{\mathbb{C}}(\pi_X)$$

Upper-bound (hard):

- Bakry–Émery theorem for log-concave π_X [Bakry and Émery 1983]

$$\mathrm{Hess}(-\log \pi_X(x)) \succeq \rho I_d \qquad \Rightarrow \qquad \overline{\mathbb{C}}(\pi_X) \leq 1/\rho$$

- Holley–Stroock perturbation lemma [Holley and Stroock 1987]

$$\alpha\, \rho_X(x) \leq \pi_X(x) \leq \beta\, \rho_X(x) \qquad \Rightarrow \qquad \overline{\mathbb{C}}(\pi_X) \leq \frac{\beta}{\alpha} C(\rho_X)$$

- Combination of the two above.

How to **control** $\overline{\mathbb{C}}(\pi_X)$? [Zahm et al, 2022]

Lower-bound (easy): by testing $\mathrm{Var}(f(X)) \leq \mathbb{C}(\pi_X)\mathbb{E}[\|\nabla f(x)\|^2]$ with affine functions, we get

$$\sup_{f:\text{affine}} \frac{\mathrm{Var}(f(X))}{\mathbb{E}[\|\nabla f(X)\|^2]} = \lambda_{\max}(\mathrm{Cov}(X)) \leq \mathbb{C}(\pi_X) \overset{\text{(definition of subspace PI)}}{\leq} \overline{\mathbb{C}}(\pi_X)$$

Upper-bound (hard):

- Bakry–Émery theorem for log-concave π_X [Bakry and Émery 1983]

$$\mathrm{Hess}(-\log \pi_X(x)) \succeq \rho I_d \qquad \Rightarrow \qquad \overline{\mathbb{C}}(\pi_X) \leq 1/\rho$$

- Holley–Stroock perturbation lemma [Holley and Stroock 1987]

$$\alpha\, \rho_X(x) \leq \pi_X(x) \leq \beta\, \rho_X(x) \qquad \Rightarrow \qquad \overline{\mathbb{C}}(\pi_X) \leq \frac{\beta}{\alpha} C(\rho_X)$$

- Combination of the two above.

Examples

- **Standard Gaussian measure** $\pi_X = \mathcal{N}(0, I_d)$: $\boxed{\overline{\mathbb{C}}(\pi_X) = 1}$ (ideal setting)
- Uniform measures on compact & convex domains
- Gaussian mixture? Yeah-ish... (worth setting)

How to **improve** $\overline{\mathbb{C}}(\pi_X)$? [Zahm et.al. 2020]

Always **whiten the input parameter**

$$\overline{X} = \mathrm{Cov}(X)^{-1/2}(X - \mathbb{E}[X])$$

before applying the subspace Poincaré inequality!! ($\Rightarrow$ dimensionless $\overline{\mathbb{C}}(\pi_{\overline{X}})$)

How to **improve** $\overline{\mathbb{C}}(\pi_X)$? [Zahm et.al. 2020]

Always **whiten the input parameter**

$$\overline{X} = \mathrm{Cov}(X)^{-1/2}(X - \mathbb{E}[X])$$

before applying the subspace Poincaré inequality!! ($\Rightarrow$ dimensionless $\overline{\mathbb{C}}(\pi_{\overline{X}})$)

$$\begin{aligned}
\mathbb{E}[(u(X) - f(U_m^\top \overline{X}))^2] &\leq \overline{\mathbb{C}}(\pi_{\overline{X}}) \Big(\mathbb{E}[\| \mathrm{Cov}(X)^{1/2} \nabla u(X) \|^2] - \mathbb{E}[\| U_m^\top \mathrm{Cov}(X)^{1/2} \nabla u(X) \|^2] \Big) \\
&= \overline{\mathbb{C}}(\pi_{\overline{X}}) \Big(\mathrm{trace}(H \mathrm{Cov}(X)) - \mathrm{trace}(U_m^\top \mathrm{Cov}(X)^{1/2} H \mathrm{Cov}(X)^{1/2} U_m) \Big)
\end{aligned}$$

How to **improve** $\overline{\mathbb{C}}(\pi_X)$? [Zahm et.al. 2020]

Always **whiten the input parameter**

$$\overline{X} = \mathrm{Cov}(X)^{-1/2}(X - \mathbb{E}[X])$$

before applying the subspace Poincaré inequality!! ($\Rightarrow$ dimensionless $\overline{\mathbb{C}}(\pi_{\overline{X}})$)

$$\begin{aligned}\mathbb{E}[(u(X) - f(U_m^\top \overline{X}))^2] &\leq \overline{\mathbb{C}}(\pi_{\overline{X}})\Big(\mathbb{E}[\|\mathrm{Cov}(X)^{1/2}\nabla u(X)\|^2] - \mathbb{E}[\|U_m^\top \mathrm{Cov}(X)^{1/2}\nabla u(X)\|^2]\Big) \\ &= \overline{\mathbb{C}}(\pi_{\overline{X}})\Big(\mathrm{trace}(H\mathrm{Cov}(X)) - \mathrm{trace}(U_m^\top \mathrm{Cov}(X)^{1/2} H \mathrm{Cov}(X)^{1/2} U_m)\Big)\end{aligned}$$

- If $X \sim \mathcal{N}(m, \Sigma)$ then $\overline{\mathbb{C}}(\pi_{\overline{X}}) = 1$
- Requires now the eigendecomposition of

$$\boxed{\overline{H} = \mathrm{Cov}(X)^{1/2}\, H\, \mathrm{Cov}(X)^{1/2}}$$

How to **improve** $\overline{\mathbb{C}}(\pi_X)$? [Zahm et.al. 2020]

Always **whiten the input parameter**

$$\overline{X} = \mathrm{Cov}(X)^{-1/2}(X - \mathbb{E}[X])$$

before applying the subspace Poincaré inequality!! ($\Rightarrow$ dimensionless $\overline{\mathbb{C}}(\pi_{\overline{X}})$)

$$\mathbb{E}[(u(X) - f(U_m^\top \overline{X}))^2] \leq \overline{\mathbb{C}}(\pi_{\overline{X}})\Big(\mathbb{E}[\|\mathrm{Cov}(X)^{1/2}\nabla u(X)\|^2] - \mathbb{E}[\|U_m^\top \mathrm{Cov}(X)^{1/2}\nabla u(X)\|^2]\Big)$$
$$= \overline{\mathbb{C}}(\pi_{\overline{X}})\Big(\mathrm{trace}(H\mathrm{Cov}(X)) - \mathrm{trace}(U_m^\top \mathrm{Cov}(X)^{1/2} H \mathrm{Cov}(X)^{1/2} U_m)\Big)$$

- If $X \sim \mathcal{N}(m, \Sigma)$ then $\overline{\mathbb{C}}(\pi_{\overline{X}}) = 1$
- Requires now the eigendecomposition of

$$\boxed{\overline{H} = \mathrm{Cov}(X)^{1/2}\, H\, \mathrm{Cov}(X)^{1/2}}$$

- A **supervised version** of the **unsupervised** $\mathrm{PCA}(X)$ (*i.e.* truncated Karhunen-Loeve expansion of X)

How to **further improve** $\overline{\mathbb{C}}(\pi_X)$? [Cui et.al. 2022]

Normalize the parameter $\overline{X} = T(X)$ with a diffeomorphic map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T(X) \sim \mathcal{N}(0, I_d)$ and apply the Poincaré inequality on $T(X)$:

$$\mathbb{E}[(u(X) - f(U_m^\top T(X)))^2] \leq \underbrace{\overline{\mathbb{C}}(\pi_{T(X)})}_{=1} \Big(\mathrm{trace}(H_T) - \mathrm{trace}(U_m^\top H_T U_m) \Big)$$

This replaces H with

$$\boxed{H_T = \mathbb{E}\left[\nabla T(X)^{-\top} \nabla u(X) \nabla u(X)^\top \nabla T(X)^{-1} \right]}$$

How to **further improve** $\overline{\mathbb{C}}(\pi_X)$? [Cui et.al. 2022]

Normalize the parameter $\overline{X} = T(X)$ with a diffeomorphic map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T(X) \sim \mathcal{N}(0, I_d)$ and apply the Poincaré inequality on $T(X)$:

$$\mathbb{E}[(u(X) - f(U_m^\top T(X)))^2] \leq \underbrace{\overline{\mathbb{C}}(\pi_{T(X)})}_{=1} \Big(\mathrm{trace}(H_T) - \mathrm{trace}(U_m^\top H_T U_m) \Big)$$

This replaces H with

$$\boxed{H_T = \mathbb{E}\left[\nabla T(X)^{-\top} \nabla u(X) \nabla u(X)^\top \nabla T(X)^{-1} \right]}$$

Towards **nonlinear** dimension reduction $u(X) \approx f(U_m^\top T(X))$

$\Rightarrow$ **Next!** $\Leftarrow$

How to **further improve** $\overline{\mathbb{C}}(\pi_X)$? [Cui et.al. 2022]

Normalize the parameter $\overline{X} = T(X)$ with a diffeomorphic map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T(X) \sim \mathcal{N}(0, I_d)$ and apply the Poincaré inequality on $T(X)$:

$$\mathbb{E}[(u(X) - f(U_m^\top T(X)))^2] \leq \underbrace{\overline{\mathbb{C}}(\pi_{T(X)})}_{=1} \Big(\mathrm{trace}(H_T) - \mathrm{trace}(U_m^\top H_T U_m) \Big)$$

This replaces H with

$$\boxed{H_T = \mathbb{E}\left[\nabla T(X)^{-\top} \nabla u(X) \nabla u(X)^\top \nabla T(X)^{-1} \right]}$$

Towards **nonlinear** dimension reduction $u(X) \approx f(U_m^\top T(X))$

$\Rightarrow$ **Next!** $\Leftarrow$

Alternatively: optimal **Riemannian metric** for Poincaré inequalities

$$\boxed{\mathrm{Var}(u(X)) \leq C(\pi_X, W) \mathbb{E}[\|\nabla u(X)\|^2_{W(X)}]}$$

What is the optimal metric $x \mapsto W(x)$ to use here? Can we use it within AS?

$\Rightarrow$ **Tomorrow!** $\Leftarrow$

(counter)Examples

An analytical (counter)example [Zahm et.al. 2020]

- Let $\pi_X = \mathcal{N}(0, I_d)$ so that $\overline{\mathbb{C}}(\pi_X) = 1$, and consider

$$u(X) \mapsto \sum_{i=1}^{d} a_i \sin(\omega_i X_i)$$

for some $a_1, \ldots a_d$ and $\omega_1, \ldots, \omega_d$.

An analytical (counter)example [Zahm et.al. 2020]

- Let $\pi_X = \mathcal{N}(0, I_d)$ so that $\overline{\mathbb{C}}(\pi_X) = 1$, and consider

$$u(X) \mapsto \sum_{i=1}^{d} a_i \sin(\omega_i X_i)$$

for some $a_1, \dots a_d$ and $\omega_1, \dots, \omega_d$.

- Restrict U_m to be a **coordinate selection** matrix so that

$$U_m^\top X = (X_{\tau_1}, \dots, X_{\tau_m})$$

for some index set $\tau \subseteq \{1, \dots, d\}$.

An analytical (counter)example [Zahm et.al. 2020]

- Let $\pi_X = \mathcal{N}(0, I_d)$ so that $\overline{\mathbb{C}}(\pi_X) = 1$, and consider

$$u(X) \mapsto \sum_{i=1}^{d} a_i \sin(\omega_i X_i)$$

for some $a_1, \dots a_d$ and $\omega_1, \dots, \omega_d$.

- Restrict U_m to be a **coordinate selection** matrix so that

$$U_m^\top X = (X_{\tau_1}, \dots, X_{\tau_m})$$

for some index set $\tau \subseteq \{1, \dots, d\}$.

- We can **compute analytically** the true error and its bound

$$\underbrace{\mathbb{E}[(u(X) - f(U_m^\top X))^2]}_{\frac{1}{2}\sum_{i \notin \tau} a_i^2(1-\exp(-2\omega_i^2))} \leq \underbrace{\overline{\mathbb{C}}(\pi_X)\Big(\mathbb{E}[\|\nabla u(X)\|^2] - \mathbb{E}[\|U_m^\top \nabla u(X)\|^2]\Big)}_{\frac{1}{2}\sum_{i \notin \tau} a_i^2 \omega_i^2(1+\exp(-2\omega_i^2))}$$

An analytical (counter)example [Zahm et.al. 2020]

$$\underbrace{\frac{1}{2}\sum_{i\notin\tau} a_i^2(1-\exp(-2\omega_i^2))}_{\left\{error\right\}} \quad \leq \quad \underbrace{\frac{1}{2}\sum_{i\notin\tau} a_i^2\omega_i^2(1+\exp(-2\omega_i^2))}_{\left\{bound\right\}}$$

- If $\omega_i = \omega$ for all $i = 1, \ldots, d$, we have

$$\arg\min_{U_m}\left\{error\right\} = \arg\min_{U_m}\left\{bound\right\}$$

An analytical (counter)example [Zahm et.al. 2020]

$$\underbrace{\frac{1}{2}\sum_{i\notin\tau} a_i^2(1-\exp(-2\omega_i^2))}_{\left\{error\right\}} \quad \leq \quad \underbrace{\frac{1}{2}\sum_{i\notin\tau} a_i^2\omega_i^2(1+\exp(-2\omega_i^2))}_{\left\{bound\right\}}$$

- If $\omega_i = \omega$ for all $i = 1, \dots, d$, we have

$$\arg\min_{U_m}\left\{error\right\} = \arg\min_{U_m}\left\{bound\right\}$$

but when $\omega \to \infty$,

$$\sqrt{\sum_{i\in-\tau} a_i^2} \quad \underset{\omega\to\infty}{\longleftarrow} \quad \min_{U_m}\left\{error\right\} \quad \leq \quad \min_{U_m}\left\{bound\right\} \quad \underset{\omega\to\infty}{\longrightarrow} \quad \infty$$

An analytical (counter)example [Zahm et.al. 2020]

$$\underbrace{\frac{1}{2}\sum_{i\notin\tau} a_i^2(1-\exp(-2\omega_i^2))}_{\left\{error\right\}} \quad \leq \quad \underbrace{\frac{1}{2}\sum_{i\notin\tau} a_i^2\omega_i^2(1+\exp(-2\omega_i^2))}_{\left\{bound\right\}}$$

- If $\omega_i = \omega$ for all $i = 1, \dots, d$, we have

$$\arg\min_{U_m}\left\{error\right\} = \arg\min_{U_m}\left\{bound\right\}$$

but when $\omega \to \infty$,

$$\sqrt{\sum_{i\in-\tau} a_i^2} \quad \underset{\omega\to\infty}{\longleftarrow} \quad \min_{U_m}\left\{error\right\} \quad \leq \quad \min_{U_m}\left\{bound\right\} \quad \underset{\omega\to\infty}{\longrightarrow} \quad \infty$$

- If $\omega_i = a_i^{-2} \geq 1$:

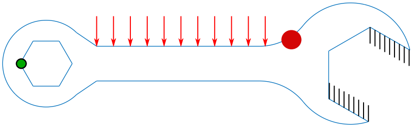
$$\arg\max_{U_m}\left\{error\right\} = \arg\min_{U_m}\left\{bound\right\}$$

aïe aïe aïe...

A numerical example: the wrenchmark

Linear elasticity problem:

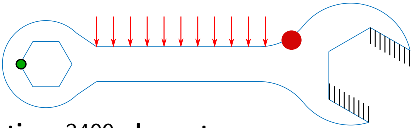
- **Input:** Young Modulus, Gaussian random field ($\overline{\mathbb{C}}(\pi_X) = 1$)
- **Output 1:** Vertical displacement of the green point
- **Output 2:** Averaged von Mises stress on the red circle



A numerical example: the wrenchmark

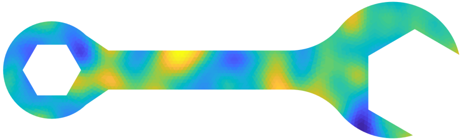
Linear elasticity problem:

- **Input:** Young Modulus, Gaussian random field ($\overline{\mathbb{C}}(\pi_X) = 1$)
- **Output 1:** Vertical displacement of the green point
- **Output 2:** Averaged von Mises stress on the red circle

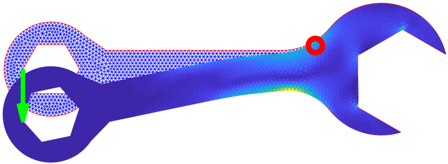


Finite element solution, 3409 **elements:**

Young modulus $= \exp(AX) \in \mathbb{R}^{3409}$,
$X \sim \mathcal{N}(0, I_d)$



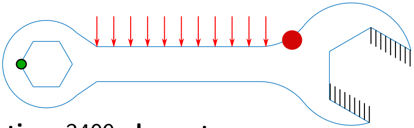
Displacement and von Mises stress



A numerical example: the wrenchmark

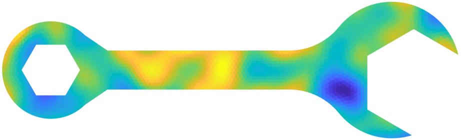
Linear elasticity problem:

- **Input:** Young Modulus, Gaussian random field ($\overline{\mathbb{C}}(\pi_X) = 1$)
- **Output 1:** Vertical displacement of the green point
- **Output 2:** Averaged von Mises stress on the red circle

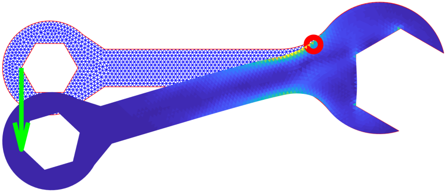


Finite element solution, 3409 **elements:**

Young modulus $= \exp(AX) \in \mathbb{R}^{3409}$,
$X \sim \mathcal{N}(0, I_d)$



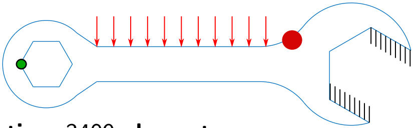
Displacement and von Mises stress



A numerical example: the wrenchmark

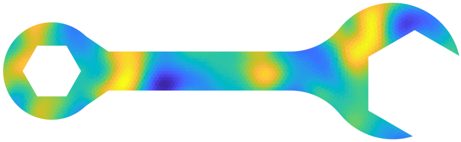
Linear elasticity problem:

- **Input:** Young Modulus, Gaussian random field ($\overline{\mathbb{C}}(\pi_X) = 1$)
- **Output 1:** Vertical displacement of the green point
- **Output 2:** Averaged von Mises stress on the red circle

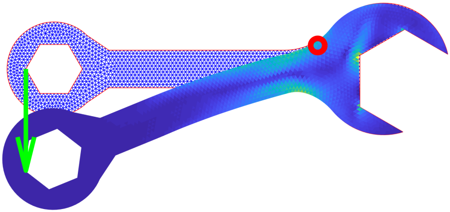


Finite element solution, 3409 **elements:**

Young modulus $= \exp(AX) \in \mathbb{R}^{3409}$,
$X \sim \mathcal{N}(0, I_d)$



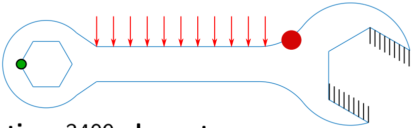
Displacement and von Mises stress



A numerical example: the wrenchmark

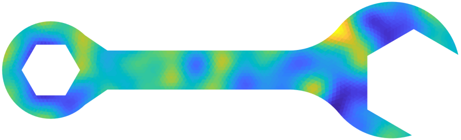
Linear elasticity problem:

- **Input:** Young Modulus, Gaussian random field ($\overline{\mathbb{C}}(\pi_X) = 1$)
- **Output 1:** Vertical displacement of the green point
- **Output 2:** Averaged von Mises stress on the red circle

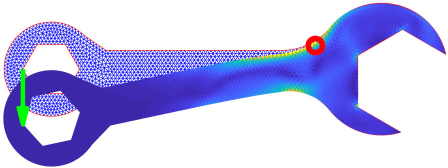


Finite element solution, 3409 **elements:**

Young modulus $= \exp(AX) \in \mathbb{R}^{3409}$,
$X \sim \mathcal{N}(0, I_d)$

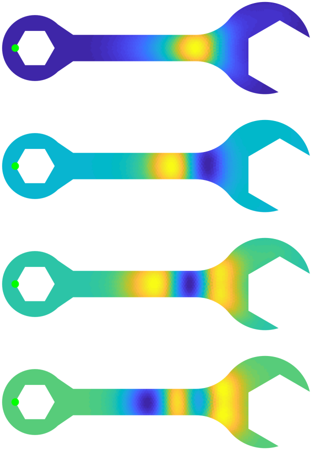


Displacement and von Mises stress



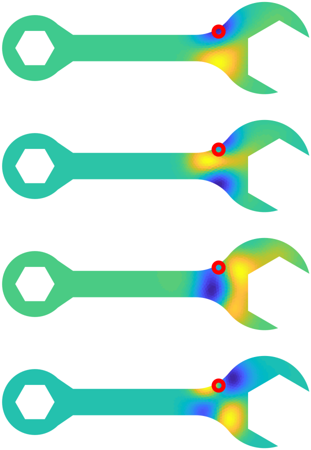
Eigenmodes $v_1, v_2, \ldots$ for $X_m = (v_1^\top X, \ldots, v_m^\top X)$

Output 1: displ. of the green point



$\ldots$

Output 2: VonMises on the red circle

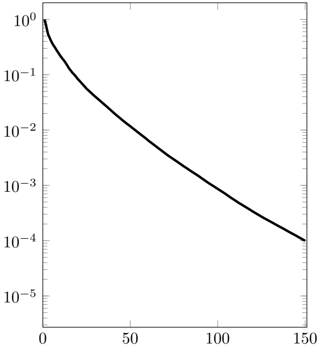


$\ldots$

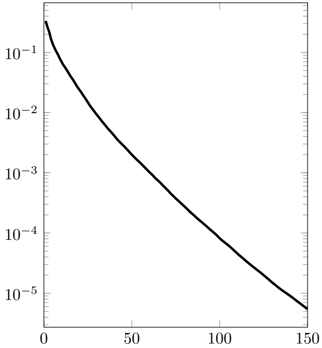
Error bound = function(m)

$$H = \mathbb{E}[\nabla u(X) \nabla u^\top(X)]$$

Output 1: displ. of the green point



Output 2: VonMises on the red circle

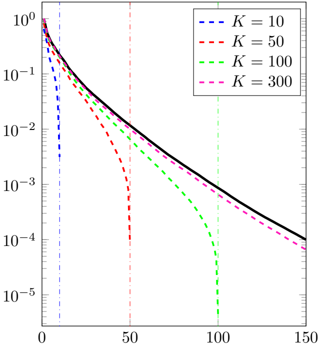


$$\{error(U_m)\} \leq \sum_{i \geq m+1} \lambda_i = function(m)$$

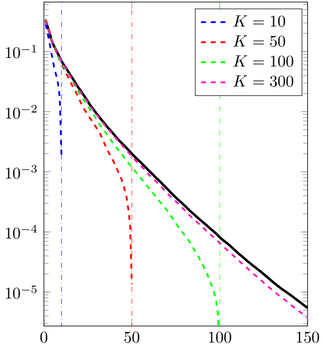
Error bound = function(m) using **MC estimation** of H

$$\hat{H} = \frac{1}{K}\sum_{i=1}^{K} \nabla u(X_i) \nabla u(X_i)^T$$

Output 1: displ. of the green point



Output 2: VonMises on the red circle

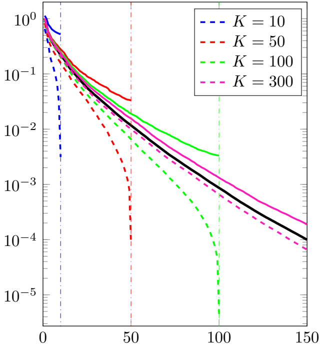


$$\{error(\hat{U}_m)\} \nleq \sum_{i \geq m+1} \hat{\lambda}_i = function(m)$$ (dashed curves)

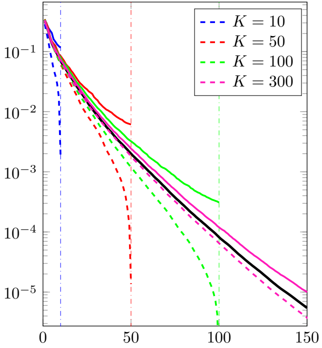
Error bound = function(m) using **MC estimation** of H

$$\hat{H} = \frac{1}{K}\sum_{i=1}^{K} \nabla u(X_i) \nabla u(X_i)^T$$

Output 1: displ. of the green point



Output 2: VonMises on the red circle



$$\{error(\hat{U}_m)\} \nleq \sum_{i \geq m+1} \hat{\lambda}_i = function(m) \qquad \text{(dashed curves)}$$

$$\{error(\hat{U}_m)\} \leq \{bound(\hat{U}_m)\} = function(m) \qquad \text{(solid curves)}$$

Linear DR **for regression**: not new, yet very active!

$$u(X_1, \dots, X_d) \approx f(U_m^\top X)$$

Dimension reduction using the gradients of the model u,

- [Samarov 1993] "Exploring **regression structure** using nonparametric functional estimation", JASA
- [Constantine et.al. 2014] "**Active subspace** methods in theory and practice", SIAM-SISC
- [Zahm et.al $\geq$ 2018] Subspace Poincaré inequalities, Vector-valued models, multi-fidelity...
- [...]

Linear DR **for regression**: not new, yet very active!

$$u(X_1, \ldots, X_d) \approx f(U_m^\top X)$$

Dimension reduction using the gradients of the model u,

- **[Samarov 1993]** "Exploring **regression structure** using nonparametric functional estimation", JASA
- **[Constantine et.al. 2014]** "**Active subspace** methods in theory and practice", SIAM-SISC
- **[Zahm et.al $\geq$ 2018]** Subspace Poincaré inequalities, Vector-valued models, multi-fidelity...
- **[...]**

or using the gradients of a surrogate model $\widetilde{u}$,

- **[Hristache et.al. 2001]** "Structure adaptive approach for dimension reduction", Ann. Stat.
- **[Fukumizu et.al. 2014]** "Gradient-based kernel dimension reduction for regression", JASA
- **[Wycoff et.al. 2021]** "Sequential learning of active subspaces", JCGS
- **[.................]**

Linear DR **for regression**: not new, yet very active!

$$u(X_1, \ldots, X_d) \approx f(U_m^\top X)$$

Dimension reduction using the gradients of the model u,

- **[Samarov 1993]** "Exploring **regression structure** using nonparametric functional estimation", JASA
- **[Constantine et.al. 2014]** "**Active subspace** methods in theory and practice", SIAM-SISC
- **[Zahm et.al $\geq$ 2018]** Subspace Poincaré inequalities, Vector-valued models, multi-fidelity...
- **[...]**

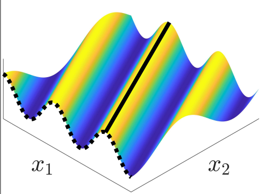
or using the gradients of a surrogate model $\widetilde{u}$,

- **[Hristache et.al. 2001]** "Structure adaptive approach for dimension reduction", Ann. Stat.
- **[Fukumizu et.al. 2014]** "Gradient-based kernel dimension reduction for regression", JASA
- **[Wycoff et.al. 2021]** "Sequential learning of active subspaces", JCGS
- **[.................]**

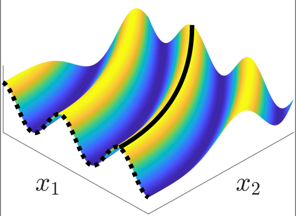
or without gradient at all!

- **[Li, 1991]** "**Sliced inverse regression** for dimension reduction", J. of the American Statistical Association.
- **[Cha, 1994]** "**Partial least squares**", Adv. Methods Mark. Res
- **[Petrushev 1998]** "Approximation by **ridge functions** and neural networks", SIAM-SIMA
- **[Cook et.al. 2005]** "**Sufficient dimension reduction** via inverse regression", JASA
- **[Fukumizu, Bach, Jordan 2004]** "Dimensionality reduction for supervised learning **with RKHS**", JMLR
- **[....]**

Towards nonlinear dimension reduction



linear



nonlinear

$$u(x) = f \circ g(x)$$

$$u(x) = f \circ g(x)$$

$$\Downarrow$$

$$\nabla u(x) = \nabla g(x)^\top \nabla f(g(x))$$

$$u(x) = f \circ g(x)$$

$$\Downarrow$$

$$\nabla u(x) = \nabla g(x)^\top \nabla f(g(x))$$

$$\Downarrow$$

$$\nabla u(x) \in \mathrm{range}(\nabla g(x)^\top)$$

$$u(x) = f \circ g(x)$$

$$\Downarrow$$

$$\nabla u(x) = \nabla g(x)^\top \nabla f(g(x))$$

$$\Downarrow$$

$$\nabla u(x) \in \mathrm{range}(\nabla g(x)^\top)$$

$$\Downarrow$$

$$J(g) := \mathbb{E}\left[\left\|\nabla u(X) - \underbrace{\Pi_{\mathrm{range}(\nabla g(X)^\top)}}_{\text{orthogonal projector}} \nabla u(X)\right\|^2\right] = 0$$

$$u(x) = f \circ g(x)$$

$$\Downarrow$$

$$\nabla u(x) = \nabla g(x)^\top \nabla f(g(x))$$

$$\Downarrow$$

$$\nabla u(x) \in \mathrm{range}(\nabla g(x)^\top)$$

$$\Downarrow$$

$$J(g) := \mathbb{E}\left[\left\| \nabla u(X) - \underbrace{\Pi_{\mathrm{range}(\nabla g(X)^\top)}}_{\text{orthogonal projector}} \nabla u(X) \right\|^2 \right] = 0$$

Two steps strategy:

1. Build g by minimizing $g \mapsto J(g)$, (aligning ∇g with ∇u)
2. Build f by minimizing $f \mapsto \mathbb{E}[(u(X) - f \circ g(X))^2]$ (low-dim regression)

$$u(x) = f \circ g(x)$$
$$\Downarrow$$
$$\nabla u(x) = \nabla g(x)^\top \nabla f(g(x))$$
$$\Downarrow$$
$$\nabla u(x) \in \mathrm{range}(\nabla g(x)^\top)$$
$$\Downarrow$$
$$J(g) := \mathbb{E}\Big[\Big\|\nabla u(X) - \underbrace{\Pi_{\mathrm{range}(\nabla g(X)^\top)}}_{\text{orthogonal projector}} \nabla u(X)\Big\|^2\Big] = 0$$

Two steps strategy:

1. Build g by minimizing $g \mapsto J(g)$, (aligning ∇g with ∇u)
2. Build f by minimizing $f \mapsto \mathbb{E}[(u(X) - f \circ g(X))^2]$ (low-dim regression)

Questions:

- **Q1:** is the reciprocal $\Uparrow$ true?
- **Q2:** does $J(g) \approx 0$ implies $u \approx f \circ g$?
- **Q3:** what approximation class $\mathcal{G}$ for g?

Question 1: is the reciprocal $\Uparrow$ true?

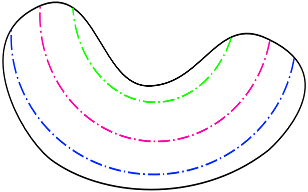
Answer: yes, under some assumptions on $\mathcal{G}$ [Bigoni et.al. 2022]

Let $g : \mathbb{R}^d \to \mathbb{R}^m$ be a smooth function such that the level-sets

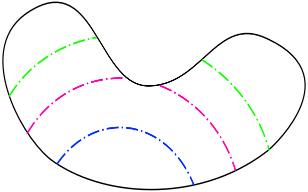
$$g^{-1}(\{z\}) = \{x \in \mathbb{R}^d : g(x) = z\},$$

are **path-connected** for any $z \in \mathbb{R}^m$. Then

$$J(g) = 0 \qquad \Rightarrow \qquad \exists f \text{ such that } u = f \circ g$$



Yes!



No...

Question 2: does $J(g) \approx 0$ implies $u \approx f \circ g$?

Answer: yes, under some assumptions on **both** $\mathcal{G}$ **and** π_X [Bigoni 2022]

Assume

$$\overline{\mathbb{C}}(X|\mathcal{G}) := \sup_{g \in \mathcal{G}} \sup_{z \in \mathbb{R}^m} \mathbb{C}(\pi_{X|g(X)=z}) < \infty$$

Then for any $g \in \mathcal{G}$, there exists a profile $f : \mathbb{R}^m \to \mathbb{R}$ such that

$$\mathbb{E}\left[\left(u(X) - f \circ g(X)\right)^2\right] \leq \overline{\mathbb{C}}(X|\mathcal{G})\, J(g)$$

Question 2: does $J(g) \approx 0$ implies $u \approx f \circ g$?

Answer: yes, under some assumptions on **both** $\mathcal{G}$ **and** π_X [Bigoni 2022]

Assume

$$\overline{\mathbb{C}}(X|\mathcal{G}) := \sup_{g \in \mathcal{G}} \sup_{z \in \mathbb{R}^m} \mathbb{C}(\pi_{X|g(X)=z}) < \infty$$

Then for any $g \in \mathcal{G}$, there exists a profile $f : \mathbb{R}^m \to \mathbb{R}$ such that

$$\mathbb{E}\left[\left(u(X) - f \circ g(X)\right)^2\right] \leq \overline{\mathbb{C}}(X|\mathcal{G})\, J(g)$$

Example: if $\mathcal{G} = \{x \mapsto U_m^\top x : U_m \in \mathbb{R}^{d \times m}$ orthorgonal columns$\}$ and if $X \sim \mathcal{N}(0, I_d)$, then

$$\overline{\mathbb{C}}(X|\mathcal{G}) = 1$$

(back to the linear dimension reduction...)

Question 3: what approximation class $\mathcal{G}$ for g?

In [Bigoni et.al. 2022], we use an **adaptive polynomial basis** $\Phi(x) = (\Phi_1(x), \dots, \Phi_p(x))$ for parametrizing g.

$$\mathcal{G} \subseteq \left\{ g(x) = G\Phi(x) : \quad G \in \mathbb{R}^{m \times p} \right\}$$

- Minimization of $g \mapsto J(g)$ via a **quasi-Newton** algorithm
- Path-connected level sets? **No...**
- Control of $\overline{\mathbb{C}}(X|\mathcal{G})$? **No...**

Question 3: what approximation class $\mathcal{G}$ for g?

In [Bigoni et.al. 2022], we use an **adaptive polynomial basis** $\Phi(x) = (\Phi_1(x), \dots, \Phi_p(x))$ for parametrizing g.

$$\mathcal{G} \subseteq \left\{ g(x) = G\Phi(x) : \quad G \in \mathbb{R}^{m \times p} \right\}$$

- Minimization of $g \mapsto J(g)$ via a **quasi-Newton** algorithm
- Path-connected level sets? **No...**
- Control of $\overline{\mathbb{C}}(X|\mathcal{G})$? **No...**

In [Verdière et.al. 2025] we employ **diffeomorphism-based** feature maps

$$\mathcal{G} \subseteq \left\{ g(x) = \begin{pmatrix} \varphi_1(x) \\ \vdots \\ \varphi_m(x) \end{pmatrix} : \quad \varphi \in \text{Diffeo}(\mathbb{R}^d; \mathbb{R}^d) \right\}$$

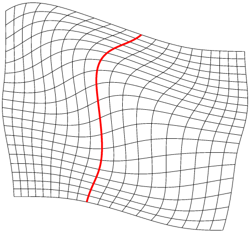
φ is parametrized using **invertible NN** (= composition of triangular monotone maps)

- Easy to minimize? **"Who cares???" (the ML answer)**
- Path-connected level set? **Yes!**
- Control of $\overline{\mathbb{C}}(X|\mathcal{G})$? **Almost..................ish**

For any diffeomorphism $\varphi : \mathbb{R}^d \to \mathbb{R}^d$, the feature map

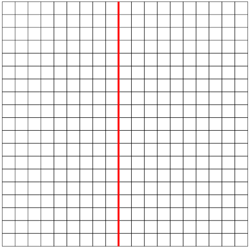
$$g(x) = \begin{pmatrix} \varphi_1(x) \\ \vdots \\ \varphi_m(x) \end{pmatrix}$$

has connected level set.



$X \in$ Input space (IS)

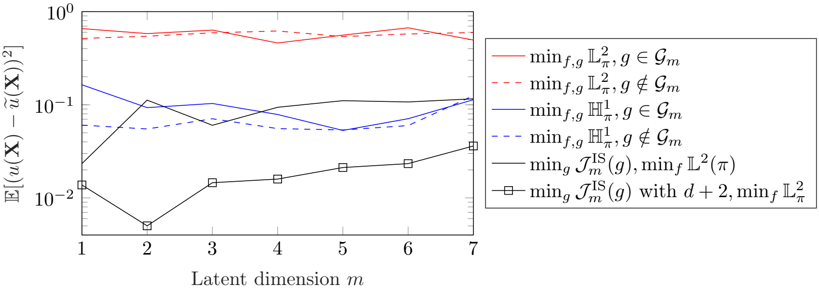
$$\xrightarrow{\varphi}$$



$\varphi(X) \in$ Feature space (FS)

Illustration

$$u(X) = \sin(\|X\|^2), \qquad X \sim \mathcal{U}([0,1]^d), \qquad d = 20$$



Using the same training set $\{X^{(i)}, u(X^{(i)}), \nabla u(X^{(i)})\}_{i=1}^N$, learn

- $g \in \mathcal{G}$ invertible NN (**solid**) or $g \notin \mathcal{G}$: fully connected NN (**dashed**)
- $f \in$ fully connected NN

by minimizing either

- $\mathbb{L}^2$-norm: $\frac{1}{N}\sum_{i=1}^N (u(X^{(i)}) - f \circ g(X^{(i)}))^2$
- $\mathbb{H}^1$-norm: $\frac{1}{N}\sum_{i=1}^N (u(X^{(i)}) - f \circ g(X^{(i)}))^2 + \|\nabla u(X^{(i)}) - \nabla(f \circ g)(X^{(i)})\|^2$
- 2 steps: first $g \in \mathcal{G}$ (via $J(g)$) then f (via the $\mathbb{L}^2$-norm)

Dimension augmentation trick [Verdière et.al. 2025]

Basic idea: approximating $u(X)$ in a dimension-augmented space via

$$\overline{u}(\underbrace{X, \Xi}_{\in \mathbb{R}^{d+k}}) = u(X)$$

where $\Xi = (\Xi_1, \ldots, \Xi_k)$ is an **arbitrary** random vector, *e.g.* $\Xi \sim \mathcal{N}(0, I_k)$

[2] [Dupont, Doucet, Teh 2019] Augmented neural odes, NeurIPS.
[3] [Huang, Dinh, Courville 2020] Augmented normalizing flows, arXiv.

Dimension augmentation trick [Verdière et.al. 2025]

Basic idea: approximating $u(X)$ in a dimension-augmented space via

$$\bar{u}(\underbrace{X, \Xi}_{\in \mathbb{R}^{d+k}}) = u(X)$$

where $\Xi = (\Xi_1, \ldots, \Xi_k)$ is an **arbitrary** random vector, *e.g.* $\Xi \sim \mathcal{N}(0, I_k)$

- As for neural ODEs[2] or normalizing flows[3], augmenting the dimension **facilitates the construction of diffeomorphisms**.

---

[2] [Dupont, Doucet, Teh 2019] Augmented neural odes, NeurIPS.

[3] [Huang, Dinh, Courville 2020] Augmented normalizing flows, arXiv.

Dimension augmentation trick [Verdière et.al. 2025]

Basic idea: approximating $u(X)$ in a dimension-augmented space via

$$\overline{u}(\underbrace{X, \Xi}_{\in \mathbb{R}^{d+k}}) = u(X)$$

where $\Xi = (\Xi_1, \ldots, \Xi_k)$ is an **arbitrary** random vector, *e.g.* $\Xi \sim \mathcal{N}(0, I_k)$

- As for neural ODEs[2] or normalizing flows[3], augmenting the dimension **facilitates the construction of diffeomorphisms**.

- For fixed $x \in \mathbb{R}$, we now approximate $u(x)$ by a **random variable**

$$u(x) \approx f \circ g(x, \Xi)$$

or eventually by its mean $u(x) \approx \mathbb{E}_{\Xi}[f \circ g(x, \Xi)]$

2 [Dupont, Doucet, Teh 2019] Augmented neural odes, NeurIPS.

3 [Huang, Dinh, Courville 2020] Augmented normalizing flows, arXiv.

Thermal block: parametrized PDE with $d = 16$

$r = 1$						
Points	Structure	#Param	MSE%	NRMSE%	RL$_2$%	RL$_1$%
$n_{\text{train}} = 100$	BACF_IS(6)	1632	$(1.25 \pm 0.59) \times 10^{-3}$	5.18 ± 1.11	12.97 ± 2.79	8.43 ± 2.77
	BACF_IS$_{d+4}$(4)	1680	$(\mathbf{6.33} \pm 1.90) \times \mathbf{10^{-4}}$	$\mathbf{3.73} \pm 0.56$	$\mathbf{9.34} \pm 1.40$	$\mathbf{5.69} \pm 1.26$
	PFM	152	$(4.76 \pm 0.27) \times 10^{-3}$	10.34 ± 0.29	25.88 ± 0.74	20.10 ± 0.96
$n_{\text{train}} = 500$	BACF_IS(6)	1632	$(1.81 \pm 0.56) \times 10^{-4}$	2.00 ± 0.31	5.01 ± 0.77	2.19 ± 0.42
	BACF_IS$_{d+4}$(4)	1680	$(\mathbf{1.32} \pm 0.55) \times \mathbf{10^{-4}}$	$\mathbf{1.68} \pm 0.35$	$\mathbf{4.22} \pm 0.87$	$\mathbf{1.92} \pm 0.44$
	PFM	152	$(4.74 \pm 0.50) \times 10^{-3}$	10.34 ± 0.55	25.90 ± 1.39	19.79 ± 0.87
$n_{\text{train}} = 2500$	NLL	-	-	6.26	-	12.71
	DRiLLS	-	-	2.19	-	2.72

$r = 2$						
Points	Structure	#Param	MSE%	NRMSE%	RL$_2$%	RL$_1$%
$n_{\text{train}} = 100$	BACF_IS(6)	1632	$(1.27 \pm 0.37) \times 10^{-3}$	5.29 ± 0.76	13.24 ± 1.91	8.28 ± 1.75
	BACF_IS$_{d+4}$(4)	1680	$(\mathbf{1.18} \pm 0.55) \times \mathbf{10^{-3}}$	$\mathbf{5.05} \pm 1.02$	$\mathbf{12.66} \pm 2.55$	$\mathbf{7.26} \pm 1.35$
	PFM	304	$(4.65 \pm 0.27) \times 10^{-3}$	10.21 ± 0.29	25.88 ± 0.73	19.78 ± 1.00
$n_{\text{train}} = 500$	BACF_IS(6)	1632	$(2.73 \pm 4.21) \times 10^{-4}$	2.18 ± 1.17	5.47 ± 2.93	3.46 ± 3.42
	BACF_IS$_{d+4}$(4)	1680	$(\mathbf{1.42} \pm 1.20) \times \mathbf{10^{-4}}$	$\mathbf{1.70} \pm 0.55$	$\mathbf{4.25} \pm 1.38$	$\mathbf{2.39} \pm 1.73$
$n_{\text{train}} = 2500$	NLL	-	-	6.66	-	13.45
	DRiLLS	-	-	-	-	-

Table: PFM is the Polynomial Feature Maps method from [Bigoni 2022] and the results for DRiLLS and NLL are copied from [Teng 2021].

Nonlinear feature learning: quite recent & super active!

$$u(X_1, \ldots, X_d) \approx f(g(X))$$

Nonlinear features using the gradients of the model u,

- **[Zhang et.a. 2019]** "Learning nonlinear level sets for DR in function approximation", NeurIPS
- **[Bigoni et.a. 2022]** "Nonlinear DR for surrogate modeling using gradient information", IMA:I&I.
- **[Romor et.a. 2022]** "Kernel-based active subspaces with application to computational fluid dynamics parametric problems using the discontinuous Galerkin method", Int. J. Numer. Methods Eng
- **[Teng et.a. 2023]** "Level Set Learning with Pseudoreversible Neural Networks for Nonlinear Dimension Reduction in Function Approximation", SIAM-SISC
- **[Nouy Pasco 2025]** "Surrogate to PI on manifolds for DR in nonlinear feature spaces", arXiv
- **[Verdière et.al 2025]** "Diffeomorphism-based feature learning using PI on augmented input space", JMLR
- **[...]**

Nonlinear feature learning: quite recent & super active!

$$u(X_1, \dots, X_d) \approx f(g(X))$$

Nonlinear features using the gradients of the model u,

- **[Zhang et.a. 2019]** "Learning nonlinear level sets for DR in function approximation", NeurIPS
- **[Bigoni et.a. 2022]** "Nonlinear DR for surrogate modeling using gradient information", IMA:I&I.
- **[Romor et.a. 2022]** "Kernel-based active subspaces with application to computational fluid dynamics parametric problems using the discontinuous Galerkin method", Int. J. Numer. Methods Eng
- **[Teng et.a. 2023]** "Level Set Learning with Pseudoreversible Neural Networks for Nonlinear Dimension Reduction in Function Approximation", SIAM-SISC
- **[Nouy Pasco 2025]** "Surrogate to PI on manifolds for DR in nonlinear feature spaces", arXiv
- **[Verdière et.al 2025]** "Diffeomorphism-based feature learning using PI on augmented input space", JMLR
- **[...]**

or without gradient at all

- **[Rosipal et.al. 2001]** "**Kernel partial least squares** regression in reproducing kernel hilbert space", JMLR
- **[Barshan et.al., 2011]** "**Supervised principal component analysis**", Pattern Recognition
- **[Lataniotis, 2020]** "Extending classical surrogate modeling to high dimensions through **supervised dimensionality reduction**", Int.JUQ
- **[.........................]**

Nonlinear feature learning: quite recent & super active!

$$u(X_1, \ldots, X_d) \approx f(g(X))$$

Nonlinear features using the gradients of the model u,

- **[Zhang et.a. 2019]** "Learning nonlinear level sets for DR in function approximation", NeurIPS
- **[Bigoni et.a. 2022]** "Nonlinear DR for surrogate modeling using gradient information", IMA:I&I.
- **[Romor et.a. 2022]** "Kernel-based active subspaces with application to computational fluid dynamics parametric problems using the discontinuous Galerkin method", Int. J. Numer. Methods Eng
- **[Teng et.a. 2023]** "Level Set Learning with Pseudoreversible Neural Networks for Nonlinear Dimension Reduction in Function Approximation", SIAM-SISC
- **[Nouy Pasco 2025]** "Surrogate to PI on manifolds for DR in nonlinear feature spaces", arXiv
- **[Verdière et.al 2025]** "Diffeomorphism-based feature learning using PI on augmented input space", JMLR
- **[...]**

or without gradient at all

- **[Rosipal et.al. 2001]** "**Kernel partial least squares** regression in reproducing kernel hilbert space", JMLR
- **[Barshan et.al., 2011]** "**Supervised principal component analysis**", Pattern Recognition
- **[Lataniotis, 2020]** "Extending classical surrogate modeling to high dimensions through **supervised dimensionality reduction**", Int.JUQ
- **[.........................]**

Unsupervised manifold learning

- **[Kernel PCA, Isomap, locally linear embedding (LLE), Hessian LLE, Laplacian eigenmaps, Linear discriminant analysis, Generalized discriminant analysis, (variational) autoencoders, Nonlinear Independent Component Analysis (ICA), Topological Data Analysis (TDA) ....]**

Questions?

Alternative loss function

$$J^{M}(g) := \mathbb{E}\Big[\big\|\nabla u(X) - \Pi^{M(X)}_{\mathrm{range}(\nabla g(X)^\top)} \nabla u(X)\big\|^2_{M(X)}\Big]$$

for some field $M(X) \succ 0$ of SPD matrices (Riemannian metric).

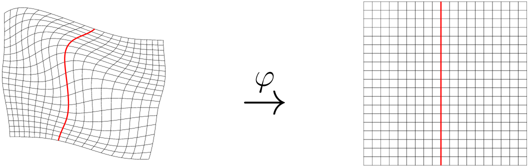
Alternative loss function

$$J^{M}(g) := \mathbb{E}\Big[\big\|\nabla u(X) - \Pi^{M(X)}_{\mathrm{range}(\nabla g(X)^\top)} \nabla u(X)\big\|^2_{M(X)}\Big]$$

for some field $M(X) \succ 0$ of SPD matrices (Riemannian metric). In particular,

$$M(X) = \nabla\varphi(X)^{-1}\nabla\varphi(X)^{-\top}$$

corresponds to **applying the Poincaré inequality in the feature space** (right).



This is equivalent to applying Active Subspace on $z \mapsto u(\varphi^{-1}(z))$

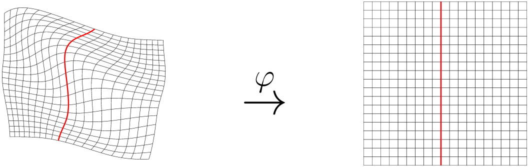
Alternative loss function

$$J^M(g) := \mathbb{E}\left[\left\|\nabla u(X) - \Pi^{M(X)}_{\mathrm{range}(\nabla g(X)^\top)} \nabla u(X)\right\|^2_{M(X)}\right]$$

for some field $M(X) \succ 0$ of SPD matrices (Riemannian metric). In particular,

$$M(X) = \nabla\varphi(X)^{-1}\nabla\varphi(X)^{-\top}$$

corresponds to **applying the Poincaré inequality in the feature space** (right).



This is equivalent to applying Active Subspace on $z \mapsto u(\varphi^{-1}(z))$

Idea: force $\varphi_\sharp \pi_X$ to being close to a Gaussian (so that $\overline{\mathbb{C}}(X|\mathcal{G}) \approx 1$?) via

$$J^M_\lambda(g) := J^M(g) + \underbrace{\lambda\, \mathrm{D}_{\mathrm{KL}}(\varphi_\sharp \pi_X || \mathcal{N}(0, I_d))}_{\text{penalization}}$$

Works in practice, but no significant improvement in our experiments...

Generalization to **vector-valued function** $u : \mathbb{R}^d \to V$

$$\mathbb{E}\Big[\underbrace{\|u(X) - f \circ g(X)\|_V^2}_{\text{vector norm}} \Big] \leq \overline{\mathbb{C}}(X|\mathcal{G})\, \mathbb{E}\Big[\underbrace{\| \overbrace{\nabla u(X)}^{\text{Jacobian}} - \Pi^{M(X)}_{\mathrm{range}(\nabla g(X)^\top)} \nabla u(X) \|^2_{V \otimes M(X)}}_{\text{Frobenius-like matrix norm}} \Big]$$

Generalization to **vector-valued function** $u : \mathbb{R}^d \to V$

$$\mathbb{E}\Big[\underbrace{\|u(X) - f \circ g(X)\|_V^2}_{\text{vector norm}}\Big] \leq \overline{\mathbb{C}}(X|\mathcal{G})\,\mathbb{E}\Big[\underbrace{\|\overbrace{\nabla u(X)}^{\text{Jacobian}} - \Pi^{M(X)}_{\mathrm{range}(\nabla g(X)^\top)} \nabla u(X)\|^2_{V \otimes M(X)}}_{\text{Frobenius-like matrix norm}}\Big]$$

Application to autoencoders: let $V = \mathbb{R}^d$ and approximate the identity function $u = id$ in $L^2(\pi_X)$

$$\begin{cases} u(X) & = X \\ \nabla u(X) & = I_d \end{cases} \quad \Leftrightarrow \quad \boxed{X \approx \underbrace{f}_{\text{decoder}} \circ \underbrace{g(X)}_{\text{encoder}}}$$

using Poincaré inequalities in the Feature Space.

Generalization to **vector-valued function** $u : \mathbb{R}^d \to V$

$$\mathbb{E}\Big[\underbrace{\| u(X) - f \circ g(X) \|_V^2}_{\text{vector norm}} \Big] \leq \overline{\mathbb{C}}(X|\mathcal{G})\, \mathbb{E}\Big[\underbrace{\| \overbrace{\nabla u(X)}^{\text{Jacobian}} - \Pi^{M(X)}_{\mathrm{range}(\nabla g(X)^\top)} \nabla u(X) \|^2_{V \otimes M(X)}}_{\text{Frobenius-like matrix norm}} \Big]$$

Application to autoencoders: let $V = \mathbb{R}^d$ and approximate the identity function $u = id$ in $L^2(\pi_X)$

$$\begin{cases} u(X) & = X \\ \nabla u(X) & = I_d \end{cases} \quad \Leftrightarrow \quad \boxed{X \approx \underbrace{f}_{\text{decoder}} \circ \underbrace{g(X)}_{\text{encoder}}}$$

using Poincaré inequalities in the Feature Space. This yields the following loss:

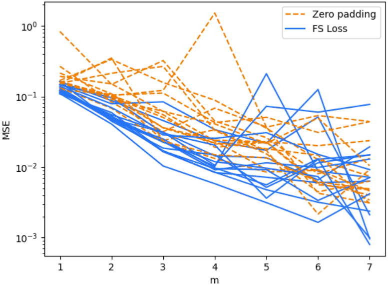
$$J^{\mathrm{AE}}_\lambda(\varphi) = \underbrace{\lambda\, \mathrm{D}_{\mathrm{KL}}(\varphi_\sharp \pi_X || \mathcal{N}(0, I_d))}_{\text{train a ``normalizing flow''...}} + \underbrace{\mathbb{E}\left[ \| U_\perp^\top \nabla \varphi(X)^{-\top} \|_F^2 \right]}_{\text{... whose last components are ``inactive''}}$$

Tests on a synthetic data set

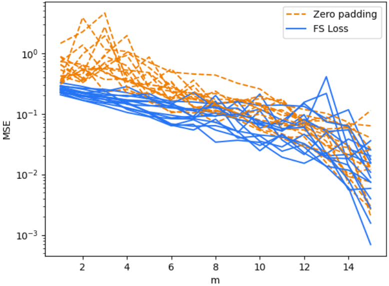
$$X = \Psi(Z) \in \mathbb{R}^d, \qquad Z \in \mathbb{R}^m$$

- $\Psi : \mathbb{R}^m \to \mathbb{R}^d$ is a arbitrary (polynomial) function.
- We used $N_{\text{train}} = 100$ samples (only)
- Compare with zero-padding strategy[4] which minimizes

$$\mathcal{L}_m(\varphi) = \mathbb{E}[\|X - \varphi^{-1}(\varphi_{\leq r}(X), 0_{d-r})\|^2]$$



$d = 8, m = 3$



$d = 16, m = 5$

[4] [Nguyen et.al. 2019] "Training invertible neural networks as autoencoders", GCPR