An Approach to Space Filling Designs in RKHS

Lorenzo Calzolari

IFPEN
Inria-Grenoble

08/10/2025
ETICS 2025, Évian-les-Bains

Thesis Information

- Email: lorenzo.calzolari@ifpen.fr
- Thesis Started: 19/11/2024
- Thesis Title: Apprentissage actif pour des entrées fonctionnelles : application à l'optimisation et à l'estimation d'ensembles admissibles
- Supervivors:
  - Clémentine Prieur, Prof., Université Grenoble Alpes
  - Céline Hélbert, MdC, EC-Lyon
  - Delphine Sinoquet, PhD, Ingénieur de recherche, IFPEN
  - Miguel Munoz Zuniga, PhD, Ingénieur de recherche, IFPEN

PART 1: FUNCTIONAL SPACE FILLING DESIGNS

Morris Maximin Criterion

Space Filling Designs consist in choosing n input points $\mathcal{D}_n = \{x_1, ..., x_n\}$, **that at cover as much as possible a domain** $\mathcal{X}$.

1. **Maximin**: solve over $\mathcal{D}_n$

$$\mathcal{D}_n^* = \text{argmax}\Phi_{Mm,n} \quad \Phi_{Mm,n}(\mathcal{D}_n) = \min_{x_i, x_{i'} \in \mathcal{D}_n} d(x_i, x_{i'});$$

2. **Morris criterion**: minimize

$$\Phi_{p,n}(\mathcal{D}_n) = \left(\sum_{i<i'} d(x_i, x_{i'})^{-p} \right)^{-\frac{1}{p}}$$

We wish to extend these techniques to RKHS.

Cloud Functions in a RKHS

Denote by $\mathcal{H}_k$ RKHS generated by a kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

- $\mathcal{H}_k = \overline{H}_k$, where

$$\forall f \in H_k \quad f = \vartheta_m^{\boldsymbol{\alpha}} = \sum_{j=1}^{m} \alpha_j k_{x_j} \quad m \in \mathbb{N}$$

  for some coefficients (**intensities**) α_j's and point (**knots**) in $\mathcal{X}$;

- With inner product

$$\langle \vartheta_m^{\boldsymbol{\alpha}}, \vartheta_{\tilde{m}}^{\boldsymbol{\beta}} \rangle_{\mathcal{H}_k} = \sum_{j=1}^{m} \sum_{j'=1}^{\tilde{m}} \alpha_j \beta_{j'} k(x_j, y_{j'}).$$

We refer to this formulation of the functions in $\mathcal{H}_k$ as **Cloud Functions**.

The Unitary Ball

- The whole RKHS is too big to cover;
- Therefore we will **draw an experimental design** of n functions over the unitary ball of the RKHS $\mathcal{H}_k$;
- **Every function in the pre-RKHS H_k is a cloud function**, we will cover the following unitary ball **consisting in functions of a fixed cloud cardinality** $m \in \mathbb{N}$

$$\mathcal{B}_m = \{f(\cdot) = \vartheta_m^{\boldsymbol{\alpha}}(\cdot) \text{ for a } m\text{-cloud}, \boldsymbol{\alpha} \in \mathbb{R}^m, \|f\|_{\mathcal{H}_k} \leq 1\};$$

Cloud Functions Designs

Using the Cloud formulations we only need to find knots and the associated intensities for the functions in our experimental design

- Identify $f_i(\cdot) = \vartheta_m^{\boldsymbol{\alpha},i}(\cdot)$ for any $i = 1, ..., n$:

$$(x_{i,1}, ..., x_{i,m}, \alpha_{i,1}, ..., \alpha_{i,m}) \leftrightarrow f_i \coloneqq \vartheta_m^{\boldsymbol{\alpha},i} = \sum_{j=1}^{m} \alpha_{i,j} k_{x_{i,j}}$$

- Setting in $\mathcal{B}_m$:

$$\mathcal{D}_{n,m}^{func} = \{f_1 = \vartheta_m^{\boldsymbol{\alpha},1}, ..., f_n = \vartheta_m^{\boldsymbol{\alpha},n}\}$$
$$\mathcal{D}_{n,m} = \{(\alpha_{i,j}, x_{i,j}) | i = 1, ..., n; j = 1, ..., m\};$$

- Then $\mathcal{D}_{n,m}^{func}$ and $\mathcal{D}_{n,m}$ **are equivalent**.

Morris Functional Criterion

Using the RKHS distance $d_{\mathcal{H}_k}$ **we can define the Functional Morris Criterion** as

$$\Phi^{func;p}(\mathcal{D}_{n,m}^{func}) = \left(\sum_{f_i, f_{i'}} d_{\mathcal{H}_k}(f_i, f_{i'})^{-p} \right)^{\frac{1}{p}}$$

Given the **equivalence** $\mathcal{D}_{n,m}^{func} \leftrightarrow \mathcal{D}_{n,m}$ we have

$$\Phi^{func;p}(\mathcal{D}_{n,m}^{func}) = \Phi_{p,n}(\mathcal{D}_{n,m}) = \left(\sum_{i<i'} d_{\mathcal{H}_k}(\vartheta_m^{\boldsymbol{\alpha},i}, \vartheta_m^{\boldsymbol{\alpha}',i'})^{-p} \right)^{\frac{1}{p}}$$

Optimization Procedure

We solve the constrained optimization problem

$$\begin{aligned} \text{minimize}_{x_{i,j},\alpha_{i,j}} \quad & \Phi^{func;p}(\mathcal{D}^{func}_{n,m}) \\ \text{subject to} \quad & \sum_{j,j'=1}^{m} \alpha_{i,j}\alpha_{i,j'}k(x_{i,j},x_{i,j'}) - 1 \leq 0 \quad i = 1,...,n; \\ & x_{i,j} \in \mathcal{X}, \quad i = 1,\ldots,n, \quad j = 1,\ldots,m \end{aligned}$$

We will apply the **Interior Point Method** from the Python library Scipy.

Dimensionality

- For the most used kernels (Gaussian, Matérn, Sobolev kernels) the corresponding RKHS **has infinite dimensions**;
- Common practice in literature consists **in truncating the Fourier expansion of a function with respect to an orthonormal basis**;
- By choosing n functions of cloud cardinality m we explore at each iteration a linear subspace of dimension nm;
- Nonetheless, any time we change a function in the design we consider a cloud of new points, hence a **different linear subspace**, so that **we keep exploring the infinite dimensional space**, and not a fixed linear subspace.

PART 2: VALIDATION AND NUMERICAL RESULTS

Validation Framework

We wish to test the performance of our method against a general **Dimension Reduction method, which consists in choosing a finite orthonormal base** $\{\psi_1, ..., \psi_M\}$ and then cover the **reduced space** $V_M = \text{Span}\{\psi_1, ..., \psi_M\}$.

- To choose these quantities we resort to the **Nystrom method**;
- Fixed the reduced dimension M, **the choice of the orthonormal basis will not influence the optimal value of the Morris criterion on the reduced space**.

The Reduced Problem

$$V_M = \text{Span}\{\hat{\psi}_1, ..., \hat{\psi}_M\} \approx \mathbb{R}^M$$

Finding a Maximin optimal design over the unitary ball of V_M ($\mathbb{B}_{V_M}(1)$) is equivalent to finding a Maximin design over the unitary ball of $\mathbb{R}^M$ ($\mathbb{B}_M(1)$).

$$\min_{\{z_1,...,z_n\}\in\mathbb{B}_M(1)} \Phi_{p,n} = \min_{\{f_1,...,f_n\}\in\mathbb{B}_{V_M}(1)} \Phi^{func}_{Mm,p}$$

1. Let $\{z_1, ..., z_n\} \in \text{argmin}\Phi_{p,M}$;
2. Set $\mathcal{E}^{Q,M}_n = \{\hat{f}_1, ..., \hat{f}_n\}$, $\hat{f}_j = \sum_{i=1}^M z_{j,i}\hat{\psi}_i$
3. $\Phi^{func}_{Mm,p}(\mathcal{E}^{Q,M}_n) = \min_{\mathbb{B}_{V_M}(1)} \Phi^{func}_{Mm,p}$.

Mercer's Theorem

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be symmetric and summable. The following **Fredholm Integral Operator** is **symmetric, self-adjoint and compact**:

$$\begin{aligned} T_k \colon L^2 &\to C(\mathcal{X}) \\ \phi &\mapsto \int_{\mathcal{X}} k(y, \cdot)\, \phi(y)\, dy. \end{aligned}$$

There exist $\{(\lambda_j, \varphi_j) | j = 1, ..., +\infty\}$ with $\lambda_1 \geq \lambda_2 \geq .... \geq 0$

$$T_k \varphi_j = \lambda_j \varphi_j \quad \langle \varphi_j, \varphi_{j'} \rangle_{L^2} = \delta_{j,j'} \quad \langle \sqrt{\lambda_j} \varphi_j, \sqrt{\lambda_{j'}} \varphi_{j'} \rangle_{\mathcal{H}_k} = \delta_{j,j'}$$

$$k(x, y) = \sum_{j=1}^{+\infty} \lambda_j \varphi_j(x) \varphi_j(y) \quad \forall x, y \in \mathcal{X},$$

the convergence of the above series being uniform for both entries of k.

Nystrom Algorithm: Setting

The Nystrom method provides a way to approximate the eigencouples of the integral operator associated to a kernel starting from a sample of points $\mathcal{D}_Q = \{x_1, ..., x_Q\}$ chosen uniformly in $\mathcal{X}$.

Define the *empirical operator* and the Gram matrix

$$(\hat{T}_k^Q f)(\cdot) = \frac{1}{Q} \sum_{q=1}^{Q} k(x_q, \cdot) f(x_q) \quad (G^Q)_{q,q'} = k(x_q, x_{q'}).$$

Denote

- $G^Q v_q^Q = \lambda_q^Q v_q^Q$ q-th eigencouple of G^Q;
- $\hat{T}_k^Q \hat{\varphi}_q = \hat{\lambda}_q \hat{\varphi}_q$ q-th eigencouple of $\hat{T}_k^Q$.

Nystrom Algorithm

The eigencouples of the empirical operator $\hat{T}_k^Q$ can be found as

$$\begin{cases} \hat{\varphi}_j(x) = \frac{\sqrt{Q}}{\lambda_j^Q} \sum_{q=1}^{Q} v_{j,q}^Q k(x_q, x) \quad j = 1, ..., Q, x \in \mathcal{X} \\ \hat{\lambda}_q = \frac{\lambda_q^Q}{Q} \\ \hat{\psi}_j(x) = \sqrt{\hat{\lambda}_q} \hat{\varphi}_j(x) = \frac{1}{\sqrt{\lambda_q^Q}} \sum_{q=1}^{Q} v_{j,q}^Q k(x_q, x) \end{cases}$$

Moreover, if $j \neq j'$ then $\langle \hat{\psi}_j, \hat{\psi}_{j'} \rangle_{\mathcal{H}_k} = \delta_{j,j'}$.

We choose $M \leq Q$ as the smallest integer ensuring that

$$\Gamma^{Q,M} = \frac{\sum_{q=1}^{M} \hat{\lambda}_{Q,q}}{\sum_{q'=1}^{Q} \hat{\lambda}_{Q,q'}} \geq 0.95.$$

Testing Procedure

- We will test the cloud functions method over $\mathcal{H}_k([0,1])$ using the **Gaussian kernel**

$$k(x,y) = exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \quad x,y \in [0,1];$$

- We tested the lengthscale parameter $\sigma = 0.01, 0.1$ and a multistart procedure of 30 starts;
- We randomly chose the starting configuration in $\mathcal{B}_m$ by uniformly choosing intensities and knots in $[0,1]$;
- **Cloud Approach**: For $n = 3, 5, 7, 10$ and $m = 1, 3, 5, 10, 20$ we find optimal cloud functions designs;
- **Nystrom based Dimension Reduction**: n as above;
- We randomly draw $Q = 250$ uniformly over $[0,1]$.

Dimensions Reduction for Gaussian Kernels

Sigma=0.1 Comparative tab: Nystrom vs Cloud Functions

index	Nystrom	1-clouds	3-clouds	5-clouds	10-clouds	20-clouds
3 functions	0.59017	0.70791	0.59018	0.59017	0.59017	0.59017
5 functions	0.66226	0.73324	0.66226	0.66226	0.66234	0.66227
7 functions	0.69575	0.74689	0.69576	0.69576	0.69576	0.69575
10 functions	0.75384	0.78774	0.75205	0.72637	0.72389	0.72389
20 functions	0.85159	1.09429	0.81302	0.79112	0.78451	0.78317

Figure: Comparison the the Functional Morris Criterion among the cloud size approach and the Reduced Space Design for $\sigma = 0.1$.

- In this case, the reduced dimension has been calculated to be $M = 7$;
- We have highlighted
  - in yellow the values for the reduced dimension;
  - in green the values in the cloud approach which are approximately the same value in Nystrom (with an error of 5×10^{-5});
  - for $n = 10, 20$ we have highlighted in blues the best computed values.

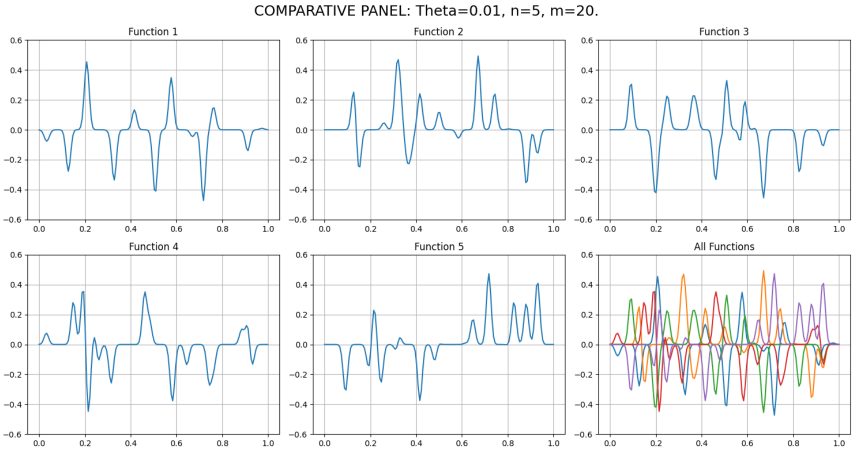

Figure: Panel with the cloud-functions generated for $\sigma = 0.01$, $n = 5$, $m = 20$

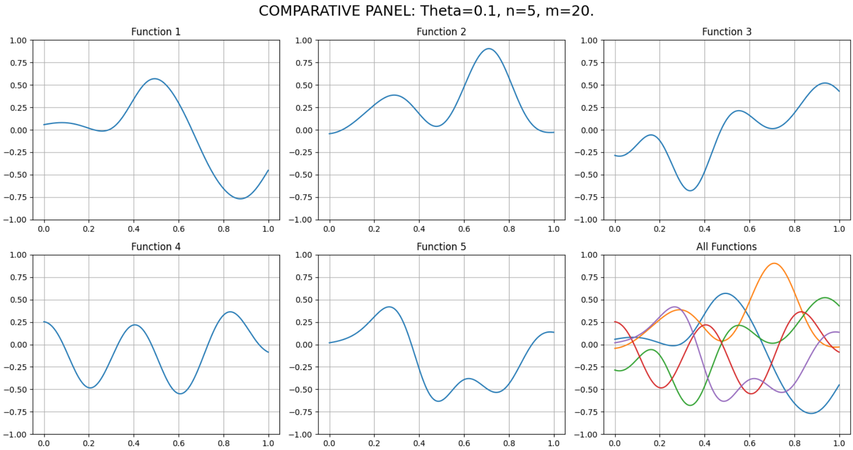

Figure: Panel with the cloud-functions generated for $\sigma = 0.1$, $n = 5$, $m = 20$

Conclusions

- The cloud based algorithm we have proposed **does not fix a finite basis** for the search of Maximin samples;
- We have shown that in terms of the functional Morris criterion, the cloud resulting designs either have similar performances ($n = 3, 5, 7$) or better performances than the dimension reduction ones ($n = 10, 20$);
- Working with Gaussian kernels:
  - σ small: very thin bell-shapes generate more irregular functions;
  - σ big: very large bell-shapes generate more regular curves;
- If we increase the cloud cardinality m the optimal Morris value stagnates.

Future Perspectives

- **Short Term Perspectives**:
  1. Deepen geometrical analysis (ongoing work);
  2. Extend numerical tests;
  3. Submit a paper hopefully by the end of the year;
- **Long Term Perspectives**:
  1. Use as initial design for functional (input) metamodelling and optimization;
  2. Application to real test cases.

THANK YOU FOR YOUR ATTENTION

Covariance Covered by Nystrom Method

To cover 95% of the covariance for the Gaussian kernel we get

- $\sigma = 0.1$: $M = 7$;
- $\sigma = 0.01$: $M = 63$.

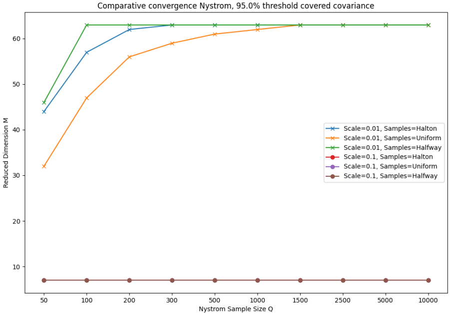

Figure: Reduced dimension M covering 95% of the covariance in for the above kernels as $Q \to +\infty$.