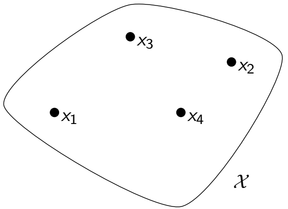
Signal Reconstruction using Determinantal Sampling

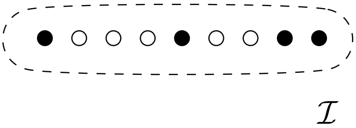
Ayoub Belhadji
OCKHAM team, LIP, ENS de Lyon
Univ Lyon, Inria, CNRS, UCBL

Joint work with **Rémi Bardenet** and **Pierre Chainais**
Centrale Lille, CRIStAL, Université de Lille, CNRS

06/12/2023 - Journée SMAI-SIGMA
LJLL, Sorbonne Université - Paris

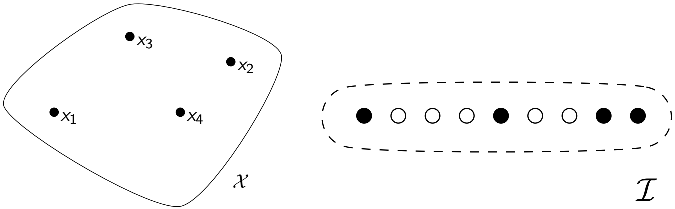
A **determinantal point process** (DPP) is a **distribution over subsets** of some set $\mathcal{X}, \mathcal{I}, \ldots$





Introduction

A **determinantal point process** (DPP) is a **distribution over subsets** of some set $\mathcal{X}, \mathcal{I}, \ldots$

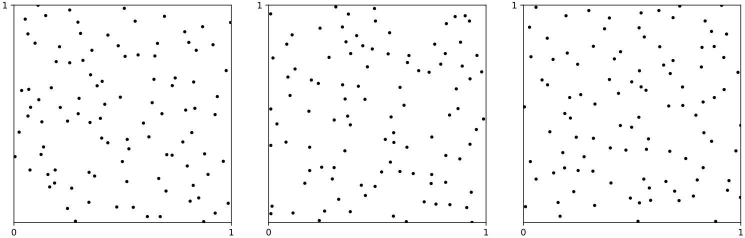


...with the **negative correlation** property:

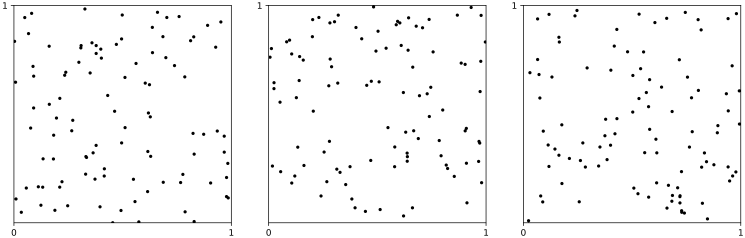
$$\forall B, B' \subset \mathcal{X}, \;\; B \cap B' = \emptyset \implies \mathbb{C}\mathrm{ov}(n_{\boldsymbol{x}}(B), n_{\boldsymbol{x}}(B')) \leq 0,$$

$$\text{where } n_{\boldsymbol{x}}(B) := |B \cap \boldsymbol{x}|$$

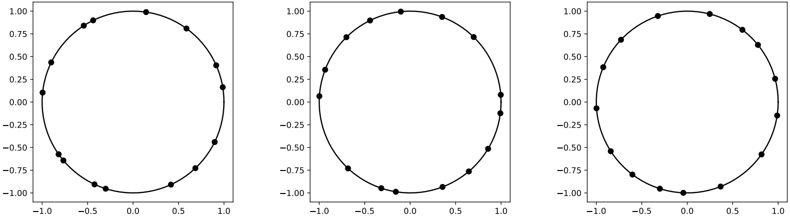
A DPP on $[0,1]^2$



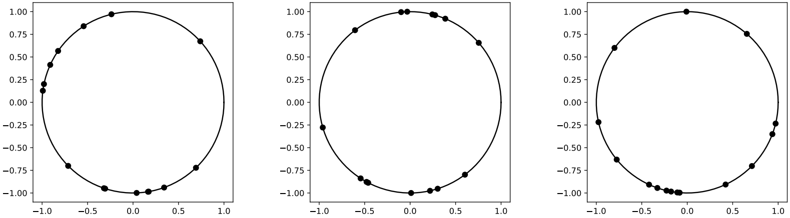
i.i.d. particles on $[0,1]^2$



A DPP on the unit circle



i.i.d. particles on the unit circle



Introduction

- Early appearances of DPPs may be traced back to the work of Dyson (1962) [1] and Ginibre (1965)[2]
- A $\underbrace{\textbf{universal}}_{\text{generic } (\mathcal{X}, \omega)}$ definition is given in the work of Macchi (1975)[3]

[1] Dyson, F.J., 1962. Statistical theory of the energy levels of complex systems. I. Journal of Mathematical Physics, 3(1), pp.140-156.

[2] Ginibre, J., 1965. Statistical ensembles of complex, quaternion, and real matrices. Journal of Mathematical Physics, 6(3), pp.440-449.

[3] Macchi, O., 1975. The coincidence approach to stochastic point processes. Advances in Applied Probability, 7(1), pp.83-122.

- Early appearances of DPPs may be traced back to the work of Dyson (1962) [1] and Ginibre (1965)[2]
- A $\underbrace{\textbf{universal}}_{\text{generic } (\mathcal{X},\omega)}$ definition is given in the work of Macchi (1975)[3]

Definition (informal): Determinantal Point Process

Given a **metric space** $\mathcal{X}$ and a **measure** ω, a DPP satisfies

$$\mathbb{P}_{\mathrm{DPP}}\left(\substack{\exists \text{ at least } k \text{ points} \\ \text{one in each } B_i,\, i=1,\dots,k}\right) = \int_{B_1\times\cdots\times B_k} \mathrm{Det}\, \underbrace{\boldsymbol{\kappa}(x_1,\dots,x_k)}_{\text{kernel matrix}} \mathrm{d}\omega(x_1)\dots\mathrm{d}\omega(x_k)$$

for a **kernel** $\kappa : \mathcal{X}\times\mathcal{X}\to\mathbb{R}$.

[1]Dyson, F.J., 1962. Statistical theory of the energy levels of complex systems. I. Journal of Mathematical Physics, 3(1), pp.140-156.

[2]Ginibre, J., 1965. Statistical ensembles of complex, quaternion, and real matrices. Journal of Mathematical Physics, 6(3), pp.440-449.

[3]Macchi, O., 1975. The coincidence approach to stochastic point processes. Advances in Applied Probability, 7(1), pp.83-122.

Introduction

Sampling using DPPs until 2019...

Discrete	Continuous
learning on budget [4]	numerical integration
node selection in a graph [5]	$\mathcal{X} \subset \mathbb{R}$ [6]
$\underbrace{\text{feature selection}}_{\text{universal results}}$ [7]	$\underbrace{\mathcal{X} = [0,1]^d}_{\text{specific results}}$ [8]

[4][Deshpande et al. (2006)],[Derezinski et al. (2017,2018,2019)]
[5][Tremblay et al. (2017)]
[6][Lambert (2018)]
[7][Belhadji et al. (2018)]
[8][Bardenet and Hardy (2016)]

Sampling using DPPs until 2019...

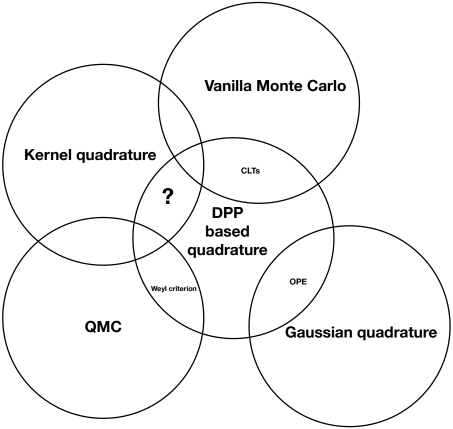
Discrete	Continuous
learning on budget [4]	numerical integration
node selection in a graph [5]	$\mathcal{X} \subset \mathbb{R}$ [6]
feature selection [7]	$\mathcal{X} = [0,1]^d$ [8]
universal results	**specific results**

universal results for
DPP-based sampling in continuous domain?

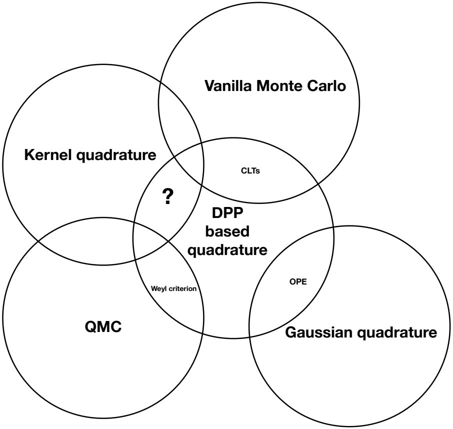
[4][Deshpande et al. (2006)],[Derezinski et al. (2017,2018,2019)]
[5][Tremblay et al. (2017)]
[6][Lambert (2018)]
[7][Belhadji et al. (2018)]
[8][Bardenet and Hardy (2016)]

Outline

Towards a universal construction of quadrature rules



Towards a universal construction of quadrature rules

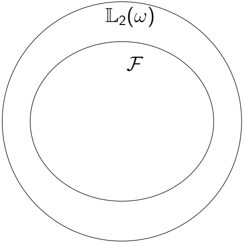


A universal construction of quadrature rules using DPPs?

Kernel-based analysis of quadrature rules

The crux of **kernel-based analysis** of a quadrature rule is the study of the **worst case error on the unit ball** of a RKHS $\mathcal{F}$

$$\sup_{\|f\|_{\mathcal{F}} \leq 1} \Big| \underbrace{\int_{\mathcal{X}} f(x) g(x) \mathrm{d}\omega(x)}_{\text{integral}} - \underbrace{\sum_{i \in [N]} w_i f(x_i)}_{\text{quadrature rule}} \Big|$$



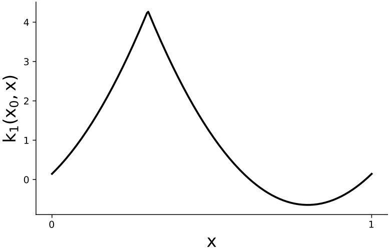
Reproducing kernel Hilbert spaces

Definition

An RKHS $\mathcal{F}$ is a Hilbert space associated to a kernel k, satisfying:

- $\forall x \in \mathcal{X},\ f \mapsto f(x)$ is continuous
- $\forall (x, f) \in \mathcal{X} \times \mathcal{F},\ \langle f, k(x, .)\rangle_{\mathcal{F}} = f(x)$

Example: $\mathcal{X} = [0, 1]$ and $k_1(x, y) := 1 - 2\pi^2 \mathcal{B}_2(\{x - y\})$ where $\{x - y\}$ is the fractional part of $x - y$, and $\mathcal{B}_2(x) = x^2 - x + \frac{1}{6}$.



Embeddings as elements of the RKHS

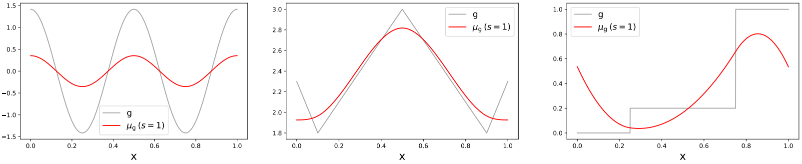
$\mathcal{F}$ contains smooth functions

Definition: an embedding of an element of $\mathbb{L}_2(\omega)$

Given $g \in \mathbb{L}_2(\omega)$, the *embedding* of g is defined by

$$\mu_g = \mathbf{\Sigma} g := \int_{\mathcal{X}} k(x,.)g(x)\mathrm{d}\omega(x) \in \mathcal{F}$$

$\mathbf{\Sigma} : \mathbb{L}_2(\omega) \to \mathbb{L}_2(\omega) =$ **integration operator** associated to (k, ω).



Embeddings and quadrature rules

Properties

- The **reproducibility of integrals**

$$\forall f \in \mathcal{F},\ \langle f, \mu_g \rangle_{\mathcal{F}} = \int_{\mathcal{X}} f(x) g(x) \mathrm{d}\omega(x).$$

- The **worst integration error on the unit ball** of $\mathcal{F}$:

$$\sup_{\|f\|_{\mathcal{F}} \leq 1} \Big| \underbrace{\int_{\mathcal{X}} f(x) g(x) \mathrm{d}\omega(x)}_{\text{integral}} - \underbrace{\sum_{i=1}^{N} w_i f(x_i)}_{\text{quadrature rule}} \Big| = \underbrace{\Big\| \mu_g - \sum_{i=1}^{N} w_i k(x_i, .) \Big\|_{\mathcal{F}}}_{\text{WCE}}$$

The study of quadrature rules boils down to the study of kernel approximations of embeddings

Embeddings and quadrature rules

A sanity check using the 'Monte Carlo quadrature'

Let $x_1, \dots, x_N = \text{i.i.d.} \sim \omega, w_i = 1/N$.
Under some assumptions on k, we have

$$\mathbb{E}\Big\|\mu_g - \sum_{i=1}^{N} \frac{1}{N} k(x_i, .)\Big\|_{\mathcal{F}}^2 = \mathcal{O}(1/N).$$

We recover the 'Monte Carlo rate' $\mathcal{O}(1/N)$

Embeddings and quadrature rules

A sanity check using the 'Monte Carlo quadrature'

Let $x_1, \dots, x_N = \text{i.i.d.} \sim \omega, w_i = 1/N$.
Under some assumptions on k, we have

$$\mathbb{E}\left\| \mu_g - \sum_{i=1}^{N} \frac{1}{N} k(x_i, .) \right\|_{\mathcal{F}}^2 = \mathcal{O}(1/N).$$

We recover the 'Monte Carlo rate' $\mathcal{O}(1/N)$

Can we do better?

The optimal kernel quadrature

Definition: the optimal kernel quadrature

Given a set of nodes $\boldsymbol{x} = \{x_1, \dots, x_N\}$ s.t.

$$\boldsymbol{K}(\boldsymbol{x}) := \begin{pmatrix} k(x_1, x_1) & \dots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \dots & k(x_N, x_N) \end{pmatrix}$$

is non-singular, the **optimal kernel quadrature** is the couple $(\boldsymbol{x}, \hat{\boldsymbol{w}})$ such that

$$\left\| \mu_g - \sum_{i=1}^{N} \hat{w}_i(\mu_g) k(x_i, .) \right\|_{\mathcal{F}} = \min_{\boldsymbol{w} \in \mathbb{R}^N} \left\| \mu_g - \sum_{i=1}^{N} w_i k(x_i, .) \right\|_{\mathcal{F}}$$

The convergence of the optimal kernel quadrature

The study of the **convergence rate** of the optimal kernel quadrature was carried out in several works

$\mathcal{X}$	$\mathcal{F}$ or k	$\boldsymbol{x}$	The rate	Reference
$[0,1]$	Sobolev S.	Unif. grid (g is cos or sin)	$\mathcal{O}(N^{-2s})$	[Novak et al., 2015] [Bojanov , 1981]
$[0,1]^d$	$\otimes$ Sobolev S.	QMC seq. (g is constant)	QMC rates	[Briol et al, 2019]
$[0,1]^d$	$\otimes$ Sobolev S.	-	$\mathcal{O}(N^{-2s/d})$	[Wendland, 2004]
$\mathbb{R}^d$	Gaussian	$\otimes$ Hermite roots (+ assumptions)	$\mathcal{O}(\exp(-\alpha N))$	[Karvonen et al., 2019]
...	...	...	...	...

Limitation

The analysis is **too specific** to the RKHS $\mathcal{F}$, to $\boldsymbol{x}$, to g...

Any universal analysis of the OKQ?

A spectral characterization of the RKHS

What is common among existing results?

Assumptions

- There exists a **spectral decomposition** $(e_m, \sigma_m)_{m \in \mathbb{N}^*}$ of $\boldsymbol{\Sigma}$ where $(e_m)_{m \in \mathbb{N}^*}$ is an o.n.b. of $\mathbb{L}_2(\omega)$ and $\sigma_1 \geq \sigma_2 \geq ... > 0$ s.t.

$$\sum_{m \in \mathbb{N}^*} \sigma_m < +\infty$$

- The RKHS $\mathcal{F}$ is **dense** in $\mathbb{L}_2(\omega)$

A spectral characterization of the RKHS

What is common among existing results?

Assumptions

- There exists a **spectral decomposition** $(e_m, \sigma_m)_{m \in \mathbb{N}^*}$ of $\boldsymbol{\Sigma}$ where $(e_m)_{m \in \mathbb{N}^*}$ is an o.n.b. of $\mathbb{L}_2(\omega)$ and $\sigma_1 \geq \sigma_2 \geq ... > 0$ s.t.

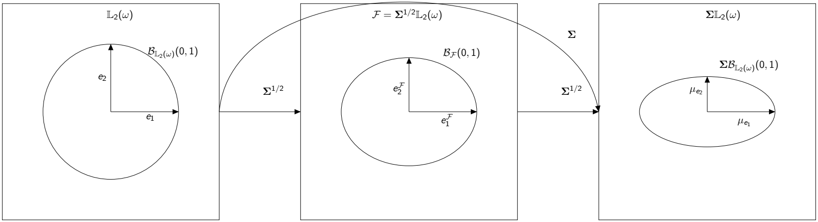
$$\sum_{m \in \mathbb{N}^*} \sigma_m < +\infty$$

- The RKHS $\mathcal{F}$ is **dense** in $\mathbb{L}_2(\omega)$

$$\|f\|_{\mathcal{F}}^2 = \sum_{m \in \mathbb{N}^*} \frac{\langle f, e_m \rangle_\omega^2}{\sigma_m} \implies \{\|f\|_{\mathcal{F}} = 1\} \textbf{ is an ellipsoid in } \mathbb{L}_2(\omega)$$

$$\sum_{m \in \mathbb{N}^*} \frac{\langle \mu_g, e_m \rangle_\omega^2}{\sigma_m^2} = \sum_{m \in \mathbb{N}^*} \langle g, e_m \rangle_\omega^2 \implies \{\mu_g; \|g\|_\omega = 1\} \textbf{ is an ellipsoid in } \mathbb{L}_2(\omega)$$

A spectral characterization of the RKHS



$\mathcal{X}$	$\mathcal{F}$ or k	σ_{N+1}	(e_m)
$[0,1]$	Sobolev	$\mathcal{O}(N^{-2s})$	Fourier
$[0,1]^d$	Korobov	$\mathcal{O}(\log(N)^{2s(d-1)}N^{-2s})$	$\otimes$ of Fourier
$[0,1]^d$	Sobolev	$\mathcal{O}(N^{-2s/d})$	"Fourier"
$\mathbb{S}^d$	Dot product	"-"	Spherical Harmonics
$\mathbb{R}$	Gaussian	$\mathcal{O}(e^{-\alpha N})$	Hermite Polys.
$\mathbb{R}^d$	Gaussian	$\mathcal{O}(e^{-\alpha d N^{1/d}})$	$\otimes$ of Hermite Polys.
...	...	...	...

The determinantal distributions

Definition

Let κ be a kernel s.t. $\int_{\mathcal{X}} \kappa(x,x)\mathrm{d}\omega(x) < +\infty$. The function

$$p_{\kappa}(x_1, \dots, x_N) \propto \operatorname{Det} \boldsymbol{\kappa}(\boldsymbol{x}) = \operatorname{Det} \begin{pmatrix} \kappa(x_1, x_1) & \dots & \kappa(x_1, x_N) \\ \vdots & \ddots & \vdots \\ \kappa(x_N, x_1) & \dots & \kappa(x_N, x_N) \end{pmatrix}$$

is a p.d.f. on $\mathcal{X}^N$ w.r.t. $\omega^{\otimes N}$.

The determinantal distributions

Definition

Let κ be a kernel s.t. $\int_{\mathcal{X}} \kappa(x,x)\mathrm{d}\omega(x) < +\infty$. The function

$$p_\kappa(x_1,\dots,x_N) \propto \operatorname{Det} \boldsymbol{\kappa}(\boldsymbol{x}) = \operatorname{Det} \begin{pmatrix} \kappa(x_1,x_1) & \dots & \kappa(x_1,x_N) \\ \vdots & \ddots & \vdots \\ \kappa(x_N,x_1) & \dots & \kappa(x_N,x_N) \end{pmatrix}$$

is a p.d.f. on $\mathcal{X}^N$ w.r.t. $\omega^{\otimes N}$.

Theorem (Belhadji et al. (2020))

For $(x_1,\dots,x_N) \sim p_\kappa$, the set $\boldsymbol{x} := \{x_1,\dots,x_N\}$ follows the distribution of a **mixture of DPPs**.

Main results

Theorem (Belhadji, Bardenet and Chainais (2019))

Under the determinantal distribution corresponding to

$$\kappa(x,y) = \mathfrak{K}(x,y) := \sum_{n \in [N]} e_n(x) e_n(y),$$

$\boldsymbol{x}$ follows the distribution of a **(projection) DPP**, and we have

$$\mathbb{E}_{p_\kappa} \sup_{\|g\|_\omega \leq 1} \left\| \mu_g - \sum_{i=1}^{N} \hat{w}_i(\mu_g) k(x_i, .) \right\|_{\mathcal{F}}^2 \leq 4N^2 r_{N+1},$$

where $r_{N+1} = \sum_{m=N+1}^{+\infty} \sigma_m$.

σ_m	Theoretical rate ($N^2 r_{N+1}$)	Empirical rate
m^{-2s}	$N^3 \mathcal{O}(\sigma_{N+1})$	$\mathcal{O}(\sigma_{N+1})$
α^m	$N^2 \mathcal{O}(\sigma_{N+1})$ $\approx \mathcal{O}(\sigma_{N+1})$	$\mathcal{O}(\sigma_{N+1})$

Main results

Theorem (Belhadji (2021))

Under the determinantal distribution corresponding to

$$\kappa(x,y) = \mathfrak{K}(x,y) := \sum_{n \in [N]} e_n(x) e_n(y),$$

$\boldsymbol{x}$ follows the distribution of a **(projection) DPP**, and we have

$$\forall g \in \mathbb{L}_2(\omega),\ \ \mathbb{E}_{p_\kappa} \left\| \mu_g - \sum_{i=1}^{N} \hat{w}_i(\mu_g) k(x_i, .) \right\|_{\mathcal{F}}^2 \leq 4 \|g\|_\omega^2 r_{N+1}$$

$$= \mathcal{O}(r_{N+1})$$

where $r_{N+1} := \sum_{m=N+1}^{+\infty} \sigma_m$.

Main results

Theorem (Belhadji, Bardenet and Chainais (2020))

Under the determinantal distribution corresponding to $\kappa = k$, we have

$$\forall g \in \mathbb{L}_2(\omega), \;\; \mathbb{E}_{p_\kappa} \left\| \mu_g - \sum_{i=1}^{N} \hat{w}_i(\mu_g) k(x_i, .) \right\|_{\mathcal{F}}^2 = \sum_{m=1}^{+\infty} \langle g, e_m \rangle_\omega^2 \epsilon_m(N)$$
$$\leq \|g\|_\omega^2 \epsilon_1(N)$$

where

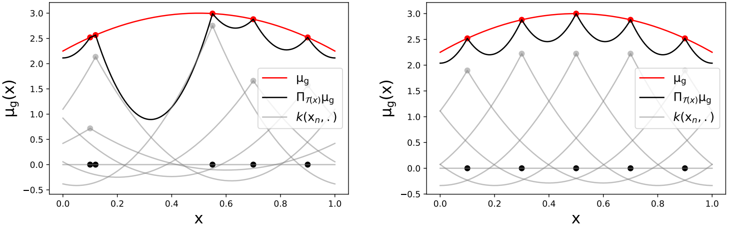
$$\epsilon_m(N) = \sum_{\substack{T \subset \mathbb{N}^* \smallsetminus \{m\} \\ |T| := N}} \prod_{t \in T} \sigma_t \Bigg/ \sum_{\substack{T \subset \mathbb{N}^* \\ |T| = N}} \prod_{t \in T} \sigma_t = \underbrace{\mathcal{O}(\sigma_{N+1})}_{\substack{\text{optimal rate} \\ \text{(Pinkus (1985))}}}$$

The optimal kernel quadrature and kernel interpolation

The mixture corresponding to OKQ is an interpolant

The function $\hat{\mu}_g := \sum_{i=1}^{N} \hat{w}_i(\mu_g)k(x_i, .)$ satisfies

$$\forall i \in [N], \ \hat{\mu}_g(x_i) = \mu_g(x_i).$$



Main results

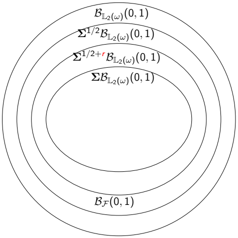
Theorem (Belhadji, Bardenet and Chainais (2020))

Under the determinantal distribution corresponding to $\kappa = k$, for $r \in [0, 1/2]$, we have

$$\forall f \in \mathbf{\Sigma}^{1/2+r}\mathbb{L}_2(\omega), \;\; \mathbb{E}_{p_\kappa} \left\| f - \sum_{i=1}^{N} \hat{w}_i(f) k(x_i, .) \right\|_{\mathcal{F}}^2 = \mathcal{O}({\sigma_{N+1}}^{2r})$$

$r = 0 \rightarrow$ a generic element of $\mathcal{F}$

$r = 1/2 \rightarrow$ an embedding of some $g \in \mathbb{L}_2(\omega)$



Main results

The RKHS norm $\|.\|_\mathcal{F}$ is strong: $\|f\|_\infty \leq \sup_{x \in \mathcal{X}} k(x,x) \|f\|_\mathcal{F}$

We seek convergence guarantees in a **weaker norm** such as $\|.\|_\omega$

Main results

Definition: least squares approximation

Given $\boldsymbol{x} = \{x_1, \dots, x_N\} \subset \mathcal{X}$, the **least squares approximation** of $f \in \mathcal{F}$ associated to $\boldsymbol{x}$ is the function $\hat{f}_{\mathrm{LS},\boldsymbol{x}}$ defined by

$$\|f - \hat{f}_{\mathrm{LS},\boldsymbol{x}}\|_\omega = \min_{\hat{f} \in \mathcal{T}(\boldsymbol{x})} \|f - \hat{f}\|_\omega,$$

where $\mathcal{T}(\boldsymbol{x}) := \mathrm{Span}(k(x_1, .), \dots, k(x_N, .))$.

Main results

Definition: least squares approximation

Given $\boldsymbol{x} = \{x_1, \dots, x_N\} \subset \mathcal{X}$, the **least squares approximation** of $f \in \mathcal{F}$ associated to $\boldsymbol{x}$ is the function $\hat{f}_{\mathrm{LS},\boldsymbol{x}}$ defined by

$$\|f - \hat{f}_{\mathrm{LS},\boldsymbol{x}}\|_\omega = \min_{\hat{f} \in \mathcal{T}(\boldsymbol{x})} \|f - \hat{f}\|_\omega,$$

where $\mathcal{T}(\boldsymbol{x}) := \mathrm{Span}(k(x_1, .), \dots, k(x_N, .))$.

$$\neq$$

Definition: optimal kernel approximation

Given $\boldsymbol{x} = \{x_1, \dots, x_N\} \subset \mathcal{X}$, the **optimal kernel approximation** of $f \in \mathcal{F}$ associated to $\boldsymbol{x}$ is the function $\hat{f}_{\mathrm{OKA},\boldsymbol{x}}$ defined by

$$\|f - \hat{f}_{\mathrm{OKA},\boldsymbol{x}}\|_{\mathcal{F}} = \min_{\hat{f} \in \mathcal{T}(\boldsymbol{x})} \|f - \hat{f}\|_{\mathcal{F}}.$$

Main results

Theorem (Belhadji, Bardenet and Chainais (2023))

Under the determinantal distribution corresponding to $\kappa = \mathfrak{K}$, we have

$$\forall f \in \mathcal{F}, \; \mathbb{E}_{p_\kappa} \|f - \hat{f}_{\mathrm{LS},\boldsymbol{x}}\|_\omega^2 \leq 2\Big(\underbrace{\|f - f_N\|_\omega^2}_{\mathcal{O}(\sigma_{N+1})} + \|f_N\|_\omega^2 \underbrace{\sum_{m=N+1}^{+\infty} \sigma_m^2}_{\substack{\mathcal{O}(\sigma_{N+1} r_{N+1}) \\ \textbf{superconvergence}}} \Big)$$

where $r_{N+1} = \sum_{m=N+1}^{+\infty} \sigma_m$.

Main results

Theorem (Belhadji, Bardenet and Chainais (2023))

Under the determinantal distribution corresponding to $\kappa = \mathfrak{K}$, we have

$$\forall f \in \mathcal{F},\ \mathbb{E}_{p_\kappa} \|f - \hat{f}_{\mathrm{LS},\boldsymbol{x}}\|_\omega^2 \leq 2\Big(\underbrace{\|f - f_N\|_\omega^2}_{\mathcal{O}(\sigma_{N+1})} + \|f_N\|_\omega^2 \underbrace{\sum_{m=N+1}^{+\infty} \sigma_m^2}_{\substack{\mathcal{O}(\sigma_{N+1} r_{N+1}) \\ \textbf{superconvergence}}} \Big)$$

where $r_{N+1} = \sum_{m=N+1}^{+\infty} \sigma_m$.

The computation of $\hat{f}_{\mathrm{LS},\boldsymbol{x}}$ requires the values of $\mu_f(x_1), \dots, \mu_f(x_N)$ not $f(x_1), \dots, f(x_N)$.

The empirical least squares approximation

Definition (Cohen, Davenport and Leviatan (2013))

Let $\boldsymbol{x} \in \mathcal{X}^N$ and $q : \mathcal{X} \to \mathbb{R}_+^*$. Consider the so-called *empirical* semi-norm $\|.\|_{q,\boldsymbol{x}}$ defined on $\mathbb{L}_2(\omega)$ by

$$\|h\|_{q,\boldsymbol{x}}^2 := \frac{1}{N} \sum_{i=1}^{N} q(x_i) h(x_i)^2.$$

The empirical least squares approximation is defined as

$$\hat{f}_{\mathrm{ELS},M,\boldsymbol{x}} := \arg\min_{\hat{f} \in \mathcal{E}_M} \|f - \hat{f}\|_{q,\boldsymbol{x}}^2,$$

where $\mathcal{E}_M := \mathrm{Span}(e_1, \ldots, e_M)$.

When $M = N$, the function $\hat{f}_{\mathrm{ELS},M,\boldsymbol{x}}$ does not depend on q.

Theorem (Belhadji, Bardenet and Chainais (2023))

Consider $N, M \in \mathbb{N}^*$ such that $M \leq N$. Let $f \in \mathcal{F}$, and define

$$\hat{f}_{\mathrm{tELS},M,\boldsymbol{x}} := \sum_{m=1}^{M} \langle \hat{f}_{\mathrm{ELS},N,\boldsymbol{x}}, e_m \rangle_\omega e_m.$$

Under the determinantal distribution corresponding to $\kappa = \mathfrak{K}$ (of cardinality N), we have

$$\mathbb{E}_{p_\kappa} \|f - \hat{f}_{\mathrm{tELS},M,\boldsymbol{x}}\|_\omega^2 = \|f - f_M\|_\omega^2 + M\|f - f_N\|_\omega^2.$$

In particular,

$$\underbrace{\mathbb{E}_{p_\kappa} \|f - \hat{f}_{\mathrm{tELS},M,\boldsymbol{x}}\|_\omega^2 \leq (1+M)\|f - f_M\|_\omega^2}_{\text{'Instance Optimal Property' (IOP)}}.$$

Some remarks:

- $\mathbb{E}_{p_\kappa}\|f - \hat{f}_{\mathrm{tELS},N,\boldsymbol{x}}\|_\omega^2 \leq (1+N)\|f - f_N\|_\omega^2 = \mathcal{O}(N\sigma_{N+1})$
- In general $\hat{f}_{\mathrm{tELS},M,\boldsymbol{x}} \neq \hat{f}_{\mathrm{ELS},M,\boldsymbol{x}}$
- The IOP was proved for $\hat{f}_{\mathrm{ELS},M,\boldsymbol{x}}$ under Christoffel sampling[9]:

$$x_1, \dots, x_N \underbrace{\sim}_{\textbf{i.i.d.}} \frac{1}{M}\sum_{m=1}^{M} e_m(x)^2 \mathrm{d}\omega(x)$$

conditioned on the event $\{\|\boldsymbol{G} - \mathbb{I}\|_{\mathrm{op}} \geq 1/2\}$, where $\boldsymbol{G} := (\langle e_i, e_j\rangle_{q,\boldsymbol{x}})_{i,j\in[M]}$ is the Gramian matrix of the family $(e_j)_{j\in[M]}$associated to the empirical scalar product $\langle .,.\rangle_{q,\boldsymbol{x}}$

[9]Cohen, A. and Migliorati, G., 2017. Optimal weighted least-squares methods. The SMAI journal of computational mathematics, 3, pp.181-203.

The determinantal distribution corresponding associated to the kernel $\kappa(x,y) := \frac{1}{M}\sum_{m=1}^{M} e_m(x)e_m(y)$ is a **natural extension of Christoffel sampling**

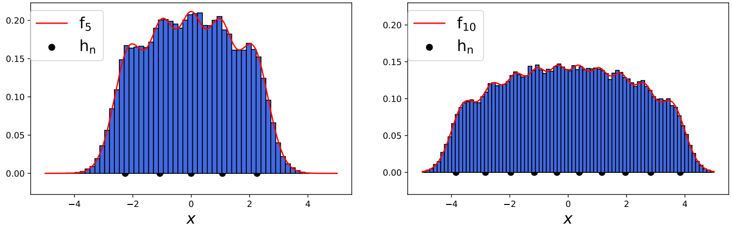
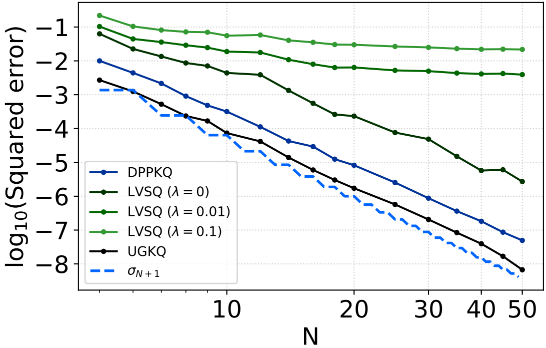


Figure: Histograms of 50000 realizations of the projection DPP associated to the first N Hermite polynomials, compared to the inverse of Christoffel function, and the nodes of the Gaussian quadrature with $N \in \{5, 10\}$

We report the empirical expectation of a surrogate of the worst interpolation error

$$\mathbb{E}_{\kappa} \sup_{\substack{f=\mu_g \\ \|g\|_\omega \leq 1}} \|f - \hat{f}_{\mathrm{OKA},\boldsymbol{x}}\|_{\mathcal{F}}^2 \approx \mathbb{E}_{\kappa} \sup_{\substack{f=\mu_g \\ g \in \mathcal{G}}} \|f - \hat{f}_{\mathrm{OKA},\boldsymbol{x}}\|_{\mathcal{F}}^2 \; ; \kappa = \mathfrak{K}$$

where $\mathcal{G} \subset \{g, \|g\|_\omega \leq 1\}$ is a finite set $|\mathcal{G}| = 5000$.
$\mathcal{F}$ is the periodic Sobolev space of order $s = 3$.



Numerical simulations

$\mathcal{F}$ = Korobov space of order $s = 1$, $\mathcal{X} = [0,1]^2$

We report $\epsilon_m(N) = \mathbb{E}_{\mathfrak{K}} \|f - \hat{f}_{\mathrm{OKA},\boldsymbol{x}}\|^2_{\mathcal{F}}; \kappa = \mathfrak{K}$ where $f = \mu_{e_m}$

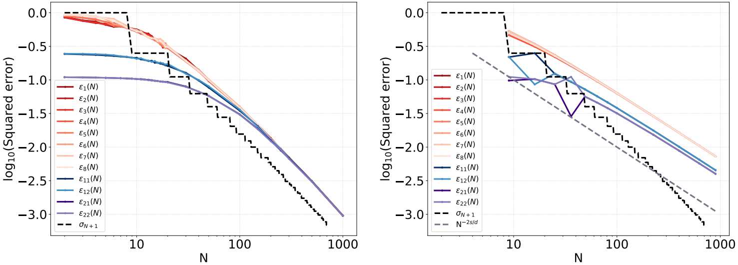


Figure: OKQ using DPPs (left) vs OKQ using the uniform grid (right)

Numerical simulations

$\mathcal{F}$ = the periodic Sobolev space of order $s = 1$, $\mathcal{X} = [0, 1]$

We report $\mathbb{E}_{\kappa}\|f - \hat{f}_{\mathrm{LS},\boldsymbol{x}}\|_{\omega}^{2}$; $\kappa = \mathfrak{K}$ where

$$f = \sum_{m=1}^{M} \xi_m e_m^{\omega}; \quad \xi_1, \ldots, \xi_M \underbrace{\sim}_{\textbf{i.i.d}} \mathcal{N}(0,1)$$

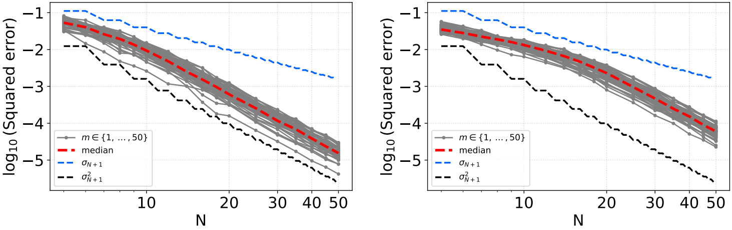


Figure: $M = 10$ (left) and $M = 20$ (right)

Numerical simulations

Consider $\mathcal{F}$ to be the RKHS defined by the Sinc kernel

$$k(x,y) = \frac{\sin(F(x-y))}{F(x-y)};\ \ \mathcal{X} = [-T/2, T/2]$$

- The eigenfunctions e_m correspond to the prolate spheroidal wave functions[10] (PSWF)
- The asymptotics of the eigenvalues in the limit $c := TF \to +\infty$ were investigated[11]: there is approximately c eigenvalues close to 1, and the remaining eigenvalues decrease to 0 at an exponential rate.

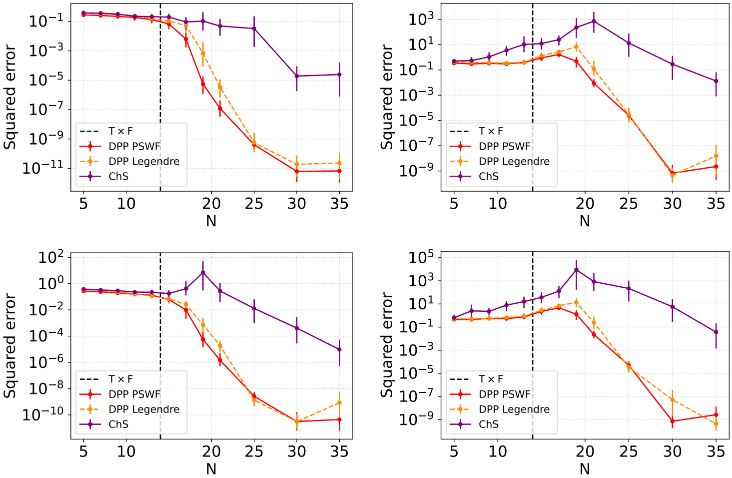
---

[10] D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, fourier analysis and uncertainty— i. Bell System Technical Journal, 40(1):43 63, 1961.

[11] H. J. Landau and H. Widom. Eigenvalue distribution of time and frequency limiting. Journal of Mathematical Analysis and Applications, 77(2):469 481, 1980.

Numerical simulations

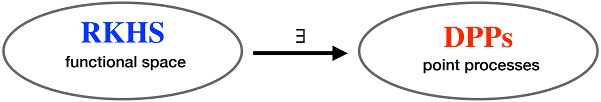
We report $\|f - \hat{f}_{\mathrm{LS},\boldsymbol{x}}\|_\omega^2$ averaged over 50 realizations for $f \in \{e_1, e_2, e_3, e_4\}$.



Conclusion

Take Home Messages

- The theoretical study of the **optimal kernel quadrature** under **determinantal sampling**
- The study of **function reconstruction** under **determinantal sampling**
- The analysis is **universal**
- **Empirical validation** on various RKHSs
- Projection DPPs are natural extensions of Christoffel sampling, and yield better empirical results



Perspectives: extending the theoretical results?

The study of

- $\mathbb{E}\|f - \hat{f}_{\mathrm{ELS},M,\boldsymbol{x}}\|_\omega^2$ when dimension $M \neq N$ number of nodes
- $\mathbb{E}\sup_{\|f\|_{\mathcal{F}}\leq 1} \|f - \hat{f}_{\mathrm{LS},\boldsymbol{x}}\|_\omega^2$ instead of $\mathbb{E}\|f - \hat{f}_{\mathrm{LS},\boldsymbol{x}}\|_\omega^2$?
- $\mathbb{E}\sup_{\|f\|_{\mathcal{F}}\leq 1} \|f - \hat{f}_{\mathrm{OKA},\boldsymbol{x}}\|_\omega^2$ and $\mathbb{E}\|f - \hat{f}_{\mathrm{OKA},\boldsymbol{x}}\|_\omega^2$?
- High order moments?
- ...

Perspectives: efficient sampling in continuous domain?

Let $\boldsymbol{x} = \{x_1, \dots, x_N\}$ such that $\operatorname{Det} \boldsymbol{\kappa}(\boldsymbol{x}) > 0$. We have

$$\begin{aligned}
\operatorname{Det} \boldsymbol{\kappa}(\boldsymbol{x}) = & \kappa(x_1, x_1) \\
& \times \left(\kappa(x_2, x_2) - \frac{\kappa(x_1, x_2)^2}{\kappa(x_1, x_1)} \right) \\
& \dots \\
& \times \left(\kappa(x_\ell, x_\ell) - \phi_{\boldsymbol{x}_{\ell-1}}(x_\ell)^T \boldsymbol{\kappa}(\boldsymbol{x}_{\ell-1})^{-1} \phi_{\boldsymbol{x}_{\ell-1}}(x_\ell) \right) \\
& \dots, \\
& \times \left(\kappa(x_N, x_N) - \phi_{\boldsymbol{x}_{N-1}}(x_N)^T \boldsymbol{\kappa}(\boldsymbol{x}_{N-1})^{-1} \phi_{\boldsymbol{x}_{N-1}}(x_N) \right)
\end{aligned}$$

where $\phi_{\boldsymbol{x}_{\ell-1}}(x) = (\kappa(\xi, x))^T_{\xi \in \boldsymbol{x}_{\ell-1}} \in \mathbb{R}^{\ell-1}$, $\boldsymbol{x}_{\ell-1} = \{x_1, \dots, x_{\ell-1}\}$.

Perspectives: efficient sampling in continuous domain?

Define

$$p_{\kappa,1}(x) = \kappa(x, x),$$
$$p_{\kappa,\ell}(x) = \kappa(x, x) - \phi_{\boldsymbol{x}_{\ell-1}}(x)^{\mathsf{T}} \boldsymbol{\kappa}(\boldsymbol{x}_{\ell-1})^{-1} \phi_{\boldsymbol{x}_{\ell-1}}(x); \;\; \ell \geq 2$$

If κ is a projection kernel

$$\int_{\mathcal{X}} p_{\kappa,\ell}(x) \mathrm{d}\omega(x) = N - \ell + 1,$$

and

$$p_{\kappa}(\boldsymbol{x}) := \frac{1}{N!} \operatorname{Det} \boldsymbol{\kappa}(\boldsymbol{x}) = \prod_{\ell=1}^{N} \frac{1}{N - \ell + 1} p_{\kappa,\ell}(x)$$

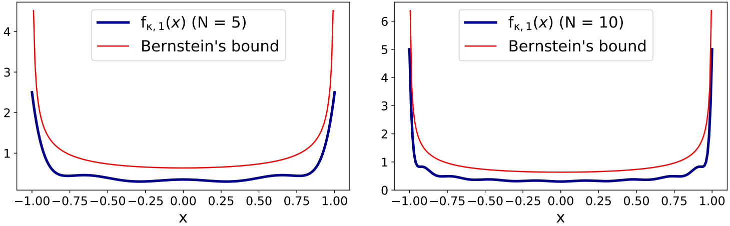
and the sequential algorithm is exact (the HKPV algorithm).

Perspectives: efficient sampling in continuous domain?

Example: to conduct Christoffel sampling for Legendre polynomials we can use the *Bernstein's bound*[12]

$$\forall n \in \mathbb{N}^*, \;\; L_n(x)^2 \leq \frac{2}{\pi}\frac{1}{\sqrt{1-x^2}} \implies p_{\kappa,1}(x) \leq \frac{2}{\pi}\frac{1}{\sqrt{1-x^2}},$$

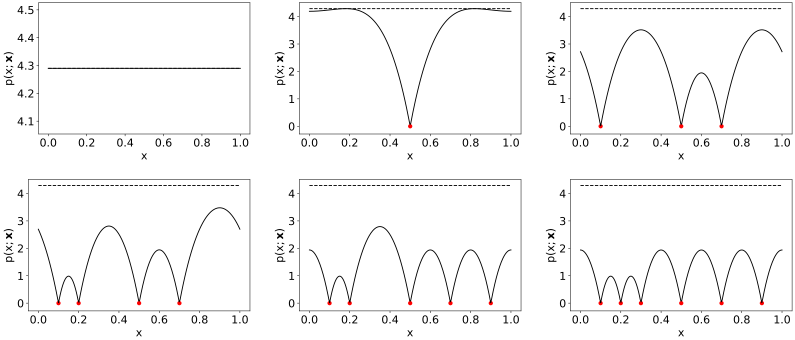
and use rejection sampling (proposal = Beta distribution).



[12]Lorch, L. (1983). "Alternative proof of a sharpened form of Bernstein's inequality for Legendre polynomials". In: Applicable Analysis 14.3, pp. 237–240.

Perspectives: efficient sampling in continuous domain?

One sample from a projection DPP (of cardinality N) requires to draw in average N^2 samples from the Christoffel distribution as a proposal: **sampling is getting harder as N get bigger**



Perspectives: efficient sampling in continuous domain?

Sampling from a projection DPP = looking for the eigenfunctions
+ looking for an upper bound for the inverse Christoffel function
+ looking for a sampling algorithm from the upper bound

How to address these issues in practice?

[13] Dolbeault, M. and Cohen, A., 2022. Optimal sampling and Christoffel functions on general domains. Constructive Approximation, 56(1), pp.121-163.

[14] Belhadji, A., Bardenet, R. and Chainais, P., 2020, November. Kernel interpolation with continuous volume sampling. In International Conference on Machine Learning (pp. 725-735). PMLR.

Perspectives: efficient sampling in continuous domain?

Sampling from a projection DPP = looking for the eigenfunctions
+ looking for an upper bound for the inverse Christoffel function
+ looking for a sampling algorithm from the upper bound

How to address these issues in practice?

We may

- Look for upper bounds or asymptotics of the inverse of Christoffel function in some RKHS on general domains [13]

[13] Dolbeault, M. and Cohen, A., 2022. Optimal sampling and Christoffel functions on general domains. Constructive Approximation, 56(1), pp.121-163.

[14] Belhadji, A., Bardenet, R. and Chainais, P., 2020, November. Kernel interpolation with continuous volume sampling. In International Conference on Machine Learning (pp. 725-735). PMLR.

Perspectives: efficient sampling in continuous domain?

Sampling from a projection DPP = looking for the eigenfunctions
+ looking for an upper bound for the inverse Christoffel function
+ looking for a sampling algorithm from the upper bound

How to address these issues in practice?

We may

- Look for upper bounds or asymptotics of the inverse of Christoffel function in some RKHS on general domains [13]
- Work with continuous volume sampling → no need to spectral decomposition[14]

[13] Dolbeault, M. and Cohen, A., 2022. Optimal sampling and Christoffel functions on general domains. Constructive Approximation, 56(1), pp.121-163.

[14] Belhadji, A., Bardenet, R. and Chainais, P., 2020, November. Kernel interpolation with continuous volume sampling. In International Conference on Machine Learning (pp. 725-735). PMLR.

Perspectives: extension to atomic measure reconstruction?

Given a class of objects $\mathcal{M}$, is it possible to approximate the elements of $\mathcal{M}$ using its evaluation on some functionals:

$$L_1(\mu), \dots, L_N(\mu) \longrightarrow \hat{\mu} \approx \mu; \;\; \mu \in \mathcal{M}$$

The objects	Functions	Atomic measures
The class $\mathcal{M}$	RKHS	not a Hilbert space
Functionals $L_1, \dots, L_N$	$\mu \mapsto \mu(x_j)$	$\mu \mapsto \int e^{\mathbf{i}\omega_j^{\mathrm{T}} x} \mathrm{d}\mu(x)$
Distance preserving P.	IOP	RIP [15]
Decoding	$\hat{f}_{\mathrm{LS},\boldsymbol{x}}, \hat{f}_{\mathrm{OKA},\boldsymbol{x}}, \dots$	CL-OMP, Mean-shift [16]

[15] Belhadji, A. and Gribonval, R., 2022. Revisiting RIP guarantees for sketching operators on mixture models.

[16] Belhadji, A. and Gribonval, R., 2022. Sketch and shift: a robust decoder for compressive clustering

Thank you for your attention!

References

A. Belhadji, R. Bardenet, and P. Chainais
Kernel quadrature with DPPs.
NeurIPS, 2019.

A. Belhadji, R. Bardenet, and P. Chainais
Kernel interpolation with continuous volume sampling
ICML, 2020.

A. Belhadji
An analysis of Ermakov-Zolotukhin quadrature using kernels.
NeurIPS, 2021.

A. Belhadji, R. Bardenet, and P. Chainais
Signal reconstruction using determinantal sampling
arXiv:2310.09437.