Kernel approximations using determinantal point processes

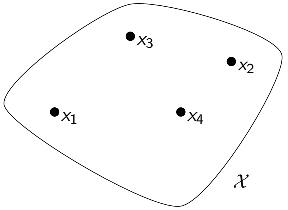
Ayoub Belhadji
ENS de Lyon

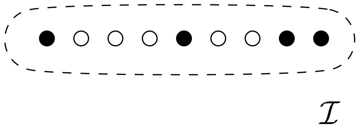
Joint work with Pierre Chainais and Rémi Bardenet

Centrale Lille, CRIStAL, Université de Lille, CNRS

Séminaire MLSP
17 mai 2022

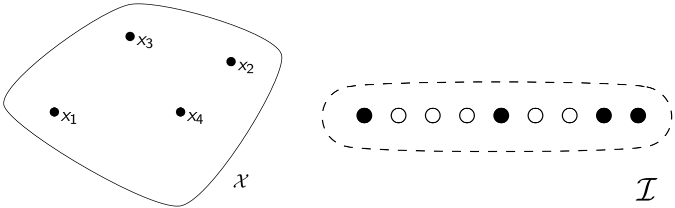
A determinantal point process (DPP) is a distribution over subsets of some set $\mathcal{X}, \mathcal{I}, \dots$ with the **negative correlation** property.





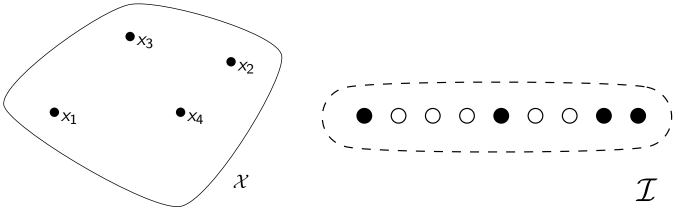
Introduction

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



A random subset $\boldsymbol{x} \subset \mathcal{X} \implies$ a **point process** $\nu_{\boldsymbol{x}} = \sum_{x \in \boldsymbol{x}} \delta_x$

A determinantal point process (DPP) is a distribution over subsets of some set $\mathcal{X}, \mathcal{I}, \dots$ with the **negative correlation** property.



A random subset $\boldsymbol{x} \subset \mathcal{X} \implies$ a **point process** $\nu_{\boldsymbol{x}} = \sum_{x \in \boldsymbol{x}} \delta_x$

characterized by the random variables $n_{\boldsymbol{x}}(B)$

$$n_{\boldsymbol{x}}(B) := |B \cap \boldsymbol{x}| = \int_{\mathcal{X}} \chi_B(x) \mathrm{d}\nu_{\boldsymbol{x}}(x)$$

where the B are Borelians.

A DPP on $\mathcal{X}$ is associated to a reference measure ω and a p.s.d. kernel κ s.t.

Introduction

A DPP on $\mathcal{X}$ is associated to a reference measure ω and a p.s.d. kernel κ s.t.

$$\mathbb{E}_{\boldsymbol{x} \sim \mathrm{DPP}}\big(n_{\boldsymbol{x}}(B)\big) = \int_B \kappa(x, x) \mathrm{d}\omega(x)$$

Introduction

A DPP on $\mathcal{X}$ is associated to a reference measure ω and a p.s.d. kernel κ s.t.

$$\mathbb{E}_{\boldsymbol{x}\sim\mathrm{DPP}}\big(n_{\boldsymbol{x}}(B)\big) = \int_B \kappa(x,x)\mathrm{d}\omega(x)$$

$$\mathbb{E}_{\boldsymbol{x}\sim\mathrm{DPP}}\big(n_{\boldsymbol{x}}(B)\times n_{\boldsymbol{x}}(B')\big) = \int_{B\times B'} \mathrm{Det}\begin{pmatrix}\kappa(x,x) & \kappa(x,x') \\ \kappa(x',x) & \kappa(x',x')\end{pmatrix}\mathrm{d}\omega(x)\mathrm{d}\omega(x')$$

A DPP on $\mathcal{X}$ is associated to a reference measure ω and a p.s.d. kernel κ s.t.

$$\mathbb{E}_{\boldsymbol{x}\sim\mathrm{DPP}}\big(n_{\boldsymbol{x}}(B)\big) = \int_B \kappa(x,x)\mathrm{d}\omega(x)$$

$$\mathbb{E}_{\boldsymbol{x}\sim\mathrm{DPP}}\big(n_{\boldsymbol{x}}(B)\times n_{\boldsymbol{x}}(B')\big) = \int_{B\times B'} \mathrm{Det}\begin{pmatrix}\kappa(x,x) & \kappa(x,x') \\ \kappa(x',x) & \kappa(x',x')\end{pmatrix} \mathrm{d}\omega(x)\mathrm{d}\omega(x')$$

In particular, it satisfies a **negative correlation** property:

$$\begin{aligned}\mathbb{C}\mathrm{ov}_{\boldsymbol{x}\sim\mathrm{DPP}}\big(n_{\boldsymbol{x}}(B), n_{\boldsymbol{x}}(B')\big) &= -\int_{B\times B'} \kappa(x,x')^2\mathrm{d}\omega(x)\mathrm{d}\omega(x') \\ &\leq 0\end{aligned}$$

Introduction

A DPP on $\mathcal{X}$ is associated to a reference measure ω and a p.s.d. kernel κ s.t.

$$\mathbb{E}_{\boldsymbol{x}\sim\mathrm{DPP}}\big(n_{\boldsymbol{x}}(B)\big) = \int_B \kappa(x,x)\mathrm{d}\omega(x)$$

$$\mathbb{E}_{\boldsymbol{x}\sim\mathrm{DPP}}\big(n_{\boldsymbol{x}}(B)\times n_{\boldsymbol{x}}(B')\big) = \int_{B\times B'} \mathrm{Det}\begin{pmatrix}\kappa(x,x) & \kappa(x,x') \\ \kappa(x',x) & \kappa(x',x')\end{pmatrix} \mathrm{d}\omega(x)\mathrm{d}\omega(x')$$

In particular, it satisfies a **negative correlation** property:

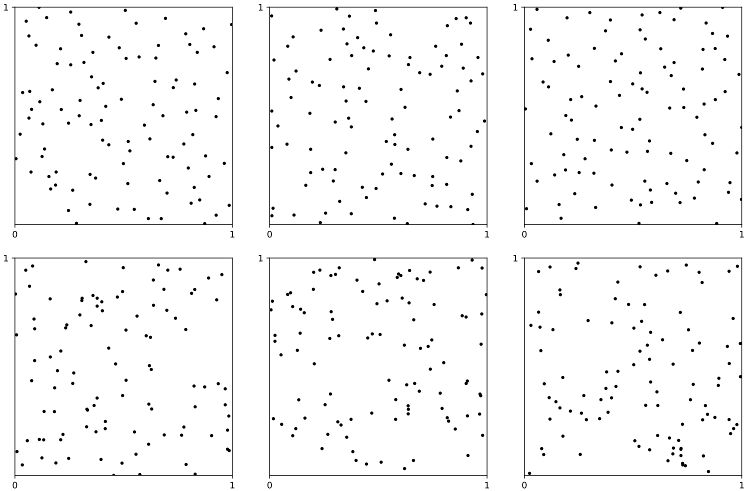
$$\begin{aligned}\mathbb{C}\mathrm{ov}_{\boldsymbol{x}\sim\mathrm{DPP}}\big(n_{\boldsymbol{x}}(B), n_{\boldsymbol{x}}(B')\big) &= -\int_{B\times B'} \kappa(x,x')^2\mathrm{d}\omega(x)\mathrm{d}\omega(x') \\ &\leq 0\end{aligned}$$

In general, we have

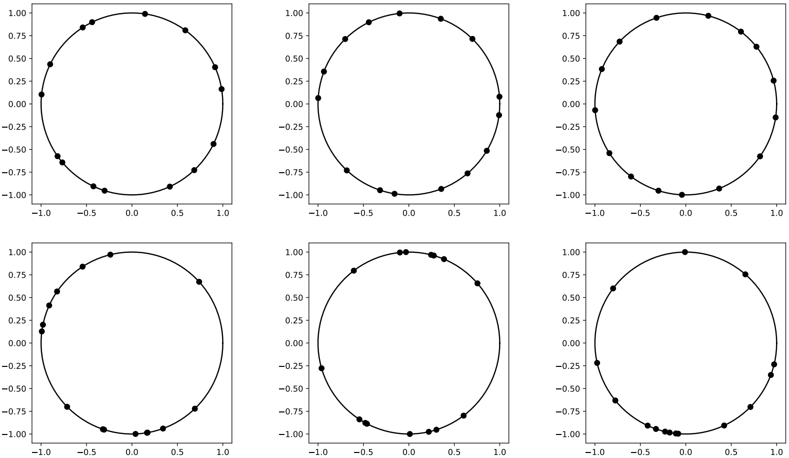
$$\mathbb{E}_{\boldsymbol{x}\sim\mathrm{DPP}}\Big(\prod_{\ell\in[L]} n_{\boldsymbol{x}}(B_\ell)\Big) = \int_{\prod_{\ell\in[L]} B_\ell} \mathrm{Det}\,\kappa(x_1,\dots,x_L) \otimes_{\ell\in[L]} \mathrm{d}\omega(x_\ell)$$

Introduction

ω is the uniform measure on $[0,1]^2$, and κ is the Dirichlet kernel



Introduction



Introduction

DPPs were used as **tools of modelisation**:

- models for (fermions in particle physics) [Macchi (1975)]
- eigenvalues of random matrices [Weyl (1946), Dyson (1962), Ginibre (1965)]
- statistical models (spatial statistics) [Lavancier et al. (2012)]

Introduction

DPPs were used as **tools of modelisation**:

- models for (fermions in particle physics) [Macchi (1975)]
- eigenvalues of random matrices [Weyl (1946), Dyson (1962), Ginibre (1965)]
- statistical models (spatial statistics) [Lavancier et al. (2012)]

they were also used as **tools of simulation**

- subset selection (feature selection, subsampling of nodes in graphs...)[Belhadji et al. (2018), Tremblay et al. (2017) ...]
- numerical integration [Bardenet and Hardy (2016)]

Introduction

Theorem (Bardenet and Hardy (2016)- Informal)

Let $x_1, \ldots, x_N \in [0,1]^d$, s.t. $\boldsymbol{x} := (x_1, \ldots, x_N)$ follows the distribution of a particular DPP and f belongs to a Sobolev space, then

$$\sqrt{N^{1+1/d}}\Big(\sum_{i \in [N]} \frac{f(x_i)}{\kappa(x_i, x_i)} - \int_{[0,1]^d} f(x)\mathrm{d}\omega(x)\Big) \xrightarrow[N \to +\infty]{law} \mathcal{N}(0, v(f)).$$

In particular

$$\mathbb{E}_{\mathrm{DPP}}\Big| \sum_{i \in [N]} \frac{f(x_i)}{\kappa(x_i, x_i)} - \int_{[0,1]^d} f(x)\mathrm{d}\omega(x)\Big|^2 = \mathcal{O}(N^{-1-1/d})$$

which is faster than Monte Carlo rate $\mathcal{O}(N^{-1})$.

A quadrature rule is an approximation scheme of an integral

$$\int_{\mathcal{X}} f(x)g(x)\mathrm{d}\omega(x) \approx \sum_{i\in[N]} w_i f(x_i).$$

A quadrature rule is an approximation scheme of an integral

$$\int_{\mathcal{X}} f(x)g(x)\mathrm{d}\omega(x) \approx \sum_{i\in[N]} w_i f(x_i).$$

Examples:

- Gaussian quadrature
- Monte Carlo method
- Quasi-Monte Carlo method
- ...

A universal quadrature rule using DPPs?

Outline

Kernel quadrature: introduction

The definition of an RKHS

The reproducing kernel Hilbert (RKHS) associated to a kernel $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is the unique Hilbert space $\mathcal{F}$ satisfying:

- reproducibility: for every $(f, x) \in \mathcal{F} \times \mathcal{X}$, $\langle f, k(x, .)\rangle_{\mathcal{F}} = f(x)$,
- continuity: for every $x \in \mathcal{X}$, $f \mapsto f(x)$ is continuous.

In the following we assume that

$$\mathcal{F} \subset \mathbb{L}_2(\omega)$$

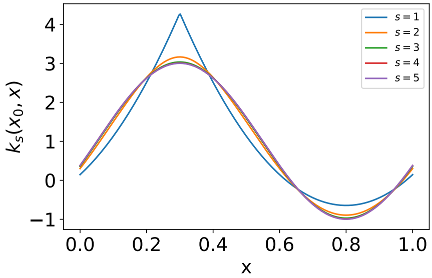
Kernel quadrature: examples

Consider $\mathcal{X} = [0,1]$, $s \in \mathbb{N}^*$ and

$$k_s(x,y) := 1 + \frac{(-1)^{s-1}(2\pi)^{2s}}{(2s)!}\mathcal{B}_{2s}(\{x-y\}),$$

where $\{x-y\}$ is the fractional part of $x-y$, and $\mathcal{B}_{2s}$ is the Bernoulli polynomial of degree $2s$:

$$\mathcal{B}_0(x) = 1,\ \mathcal{B}_2(x) = x^2 - x + \frac{1}{6},\ \mathcal{B}_4(x) = x^4 - 2x^3 + x^2 - \frac{1}{30}\dots$$



Kernel quadrature: examples

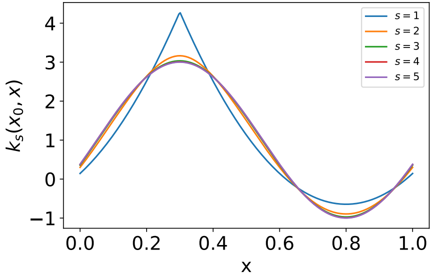
The corresponding RKHS is the periodic Sobolev space of order s:

$$\mathcal{F}_s = \left\{ f \in \mathbb{L}_2([0,1]), f(0) = f(1),\ f, f^{'}, \dots, f^{(s)} \in \mathbb{L}_2([0,1]) \right\},$$

and the corresponding norm is the Sobolev norm:

$$\|f\|^2_{\mathcal{F}_s} = \left| \int_0^1 f(x)\mathrm{d}x \right|^2 + \sum_{m \in \mathbb{N}^*} m^{2s} \left| \int_0^1 f(x) e^{-2\pi i m x} \mathrm{d}x \right|^2.$$

$$\dots \subset \mathcal{F}_4 \subset \mathcal{F}_3 \subset \mathcal{F}_2 \subset \mathcal{F}_1$$

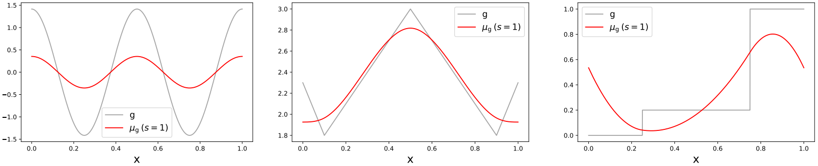


Kernel quadrature: the embeddings

The embeddings

Let $g \in \mathbb{L}_2(\omega)$, the *mean element* of $g\mathrm{d}\omega$, or the *embedding* of g is defined by

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



In particular, we have

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

Kernel quadrature: the embeddings

Assume that $f \in \mathcal{F}$. We have

$$\Big| \int_{\mathcal{X}} f(x)g(x)\mathrm{d}\omega(x) - \sum_{i\in[N]} w_i f(x_i) \Big| = \big|\langle f, \mu_g - \sum_{i\in[N]} w_i k(x_i,.)\rangle_{\mathcal{F}}\big|,$$

$$\leq \|f\|_{\mathcal{F}} \|\mu_g - \sum_{i\in[N]} w_i k(x_i,.)\|_{\mathcal{F}}.$$

Hickernell proposed to use the following figure of merite

The worst integration error on the unit ball

$$\Big\|\mu_g - \sum_{i\in[N]} w_i k(x_i,.)\Big\|_{\mathcal{F}} = \sup_{\|f\|_{\mathcal{F}}\leq 1} \Big| \int_{\mathcal{X}} f(x)g(x)\mathrm{d}\omega(x) - \sum_{i\in[N]} w_i f(x_i)\Big|$$

Kernel quadrature: the worst integration error

Observe that

$$\left\| \mu_g - \sum_{i \in [N]} w_i k(x_i, .) \right\|_{\mathcal{F}}^2 = \|\mu_g\|_{\mathcal{F}}^2 - 2\boldsymbol{w}^{\mathrm{T}} \boldsymbol{\mu_g}(\boldsymbol{x}) + \boldsymbol{w}^{\mathrm{T}} \boldsymbol{K}(\boldsymbol{x}) \boldsymbol{w}$$

where

- $\boldsymbol{\mu}_g(\boldsymbol{x}) = (\mu_g(x_i))_{i \in [N]} \in \mathbb{R}^N$
- $\boldsymbol{K}(\boldsymbol{x}) = (k(x_{i_1}, x_{i_2}))_{i_1, i_2 \in [N]} \in \mathbb{R}^{N \times N}$

Kernel quadrature: Monte Carlo approximation

Theorem (Berlinet and Thomas-Agnan (2004))

Under some assumptions, if we take $x_1, \ldots, x_N$ to be i.i.d. particles $\sim \omega$, then we have

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

Can we improve on the rate $\mathcal{O}(1/N)$?

Kernel quadrature: optimal kernel quadrature

Definition

Given a set of nodes $\boldsymbol{x} = \{x_1, \dots, x_N\}$ s.t. $\boldsymbol{K}(\boldsymbol{x})$ is non-singular, the **optimal kernel quadrature** is the couple $(\boldsymbol{x}, \hat{\boldsymbol{w}})$ such that

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

In particular

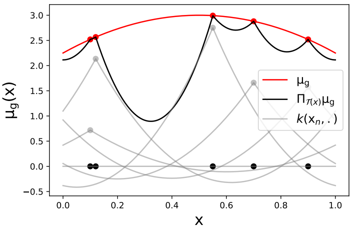
$$\left\| \mu_g - \sum_{i \in [N]} \hat{w}_i k(x_i, .) \right\|_{\mathcal{F}} = \left\| \mu_g - \boldsymbol{\Pi}_{\mathcal{T}(\boldsymbol{x})} \mu_g \right\|_{\mathcal{F}},$$

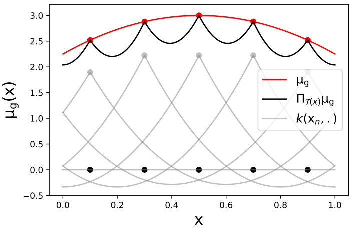
$\boldsymbol{\Pi}_{\mathcal{T}(\boldsymbol{x})}$: the orthogonal projection onto $\mathcal{T}(\boldsymbol{x}) = \mathrm{Span}(k(x_i, .))_{i \in [N]}$.

Kernel interpolation

The optimal mixture $\hat{\mu}_g := \sum_{i \in [N]} \hat{w}_i k(x_i, .)$ satisfies

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





Kernel quadrature: optimal kernel quadrature

$\mathcal{X}$	$\mathcal{F}$ 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]
$\mathbb{R}^d$	Gaussian	$\otimes$ Hermite roots ($+$ assumptions)	$\mathcal{O}(\exp(-\alpha N))$	[Karvonen et al., 2019]
Generic	Generic	?	?	-

Limitation

This kind of analysis is too specific to the RKHS $\mathcal{F}$, to $\boldsymbol{x}$, to g...

Kernel quadrature: optimal kernel quadrature

$\mathcal{X}$	$\mathcal{F}$ 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]
$\mathbb{R}^d$	Gaussian	$\otimes$ Hermite roots ($+$ assumptions)	$\mathcal{O}(\exp(-\alpha N))$	[Karvonen et al., 2019]
Generic	Generic	?	?	-

Limitation

This kind of analysis is too specific to the RKHS $\mathcal{F}$, to $\boldsymbol{x}$, to g...

An alternative analysis was proposed in Bach (2017)

Kernel quadrature: the spectral decomposition

Let $\boldsymbol{\Sigma} : \mathbb{L}_2(\omega) \rightarrow \mathbb{L}_2(\omega)$ be the integration operator

$$\boldsymbol{\Sigma} g(.) = \int_{\mathcal{X}} g(x) k(x,.) \mathrm{d}\omega(x).$$

Spectral theorem

There exist 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.

$$\boldsymbol{\Sigma} = \sum_{m \in \mathbb{N}^*} \sigma_m e_m \otimes e_m$$

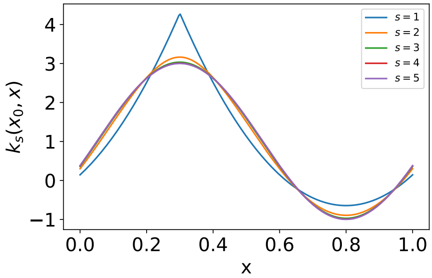
$$\Longrightarrow \quad \boldsymbol{\Sigma} g = \sum_{m \in \mathbb{N}^*} \sigma_m \langle g, e_m \rangle_{\omega} e_m$$

Kernel quadrature: an example

The kernel k_s satisfies the following identity [Wahba 90]

$$k_s(x, y) = 1 + \sum_{m \in \mathbb{N}^*} \frac{1}{m^{2s}} \cos(2\pi m(x - y))$$

it is equivalent to the Mercer decomposition with $\sigma_m = \mathcal{O}(m^{-2s})$ and $(e_m)_{m \in \mathbb{N}^*}$ is the Fourier family



Kernel quadrature: the spectral decomposition

The spectral characterization of the RKHS and the kernel

When $\mathcal{F}$ is dense in $\mathbb{L}_2(\omega)$, $(e_m^{\mathcal{F}})_{m \in \mathbb{N}^*}$ is an o.n.b. of $\mathcal{F}$, where

$$e_m^{\mathcal{F}} := \sqrt{\sigma_m} e_m,$$

so that $\langle f, e_m^{\mathcal{F}} \rangle_{\mathcal{F}} = \langle f, e_m \rangle_{\omega} / \sqrt{\sigma_m}$.

Kernel quadrature: the spectral decomposition

The spectral characterization of the RKHS and the kernel

When $\mathcal{F}$ is dense in $\mathbb{L}_2(\omega)$, $(e_m^{\mathcal{F}})_{m\in\mathbb{N}^*}$ is an o.n.b. of $\mathcal{F}$, where

$$e_m^{\mathcal{F}} := \sqrt{\sigma_m} e_m,$$

so that $\langle f, e_m^{\mathcal{F}} \rangle_{\mathcal{F}} = \langle f, e_m \rangle_{\omega} / \sqrt{\sigma_m}$. In particular

$$\|f\|_{\mathcal{F}}^2 = \sum_{m\in\mathbb{N}^*} \langle f, e_m \rangle_{\omega}^2 / \sigma_m < +\infty$$

Kernel quadrature: the spectral decomposition

The spectral characterization of the RKHS and the kernel

When $\mathcal{F}$ is dense in $\mathbb{L}_2(\omega)$, $(e_m^{\mathcal{F}})_{m\in\mathbb{N}^*}$ is an o.n.b. of $\mathcal{F}$, where

$$e_m^{\mathcal{F}} := \sqrt{\sigma_m} e_m,$$

so that $\langle f, e_m^{\mathcal{F}} \rangle_{\mathcal{F}} = \langle f, e_m \rangle_\omega / \sqrt{\sigma_m}$. In particular

$$\|f\|_{\mathcal{F}}^2 = \sum_{m\in\mathbb{N}^*} \langle f, e_m \rangle_\omega^2 / \sigma_m < +\infty$$

Moreover, we have (the Mercer decomposition)

$$k(x, y) = \sum_{m\in\mathbb{N}^*} \sigma_m e_n(x) e_n(y)$$

Kernel quadrature: ridge leverage scores sampling

Bach (2017) proposed the following quadrature:

- the x_i are sampled i.i.d. from some proposal distribution q,
- the vector of weights $\boldsymbol{w}_q(\lambda)$ solves the optimization problem

$$\min_{\boldsymbol{w} \in \mathbb{R}^N} \left\| \mu_g - \sum_{i \in [N]} \frac{w_i}{q(x_i)^{1/2}} k(x_i, .) \right\|_{\mathcal{F}}^2 + \lambda N \|\boldsymbol{w}\|_2^2,$$

for some regularization parameter $\lambda > 0$.

Kernel quadrature: ridge leverage scores sampling

Theorem (Bach 2017)

Let

$$q_\lambda(x) = \sum_{m \in \mathbb{N}^*} \frac{\sigma_m}{\sigma_m + \lambda} e_m(x)^2,$$

and $d_{\mathrm{eff}}(\lambda) = \sum_{m \in \mathbb{N}^*} \sigma_m / (\sigma_m + \lambda)$. *Assume that*

$$N \geq 5 d_{\mathrm{eff}}(\lambda) \log(16 d_{\mathrm{eff}}(\lambda)/\delta),$$

then

$$\mathbb{P}\left(\sup_{\|g\|_\omega \leq 1} \left\| \mu_g - \sum_{i \in [N]} \tilde{w}_i^q(\lambda) k(x_i, .) \right\|_{\mathcal{F}}^2 \leq 4\lambda \right) \geq 1 - \delta,$$

where $\tilde{w}_i(\lambda) = w_i(\lambda)/q_\lambda(x_i)^{1/2}$.

Kernel quadrature: ridge leverage scores sampling

In practice, in many cases

$$d_{\mathrm{eff}}(\sigma_N) \approx N, \ \ (\lambda = \sigma_N)$$

so that (up to logarithmic terms)

$$\sup_{\|g\|_\omega \leq 1} \left\| \mu_g - \sum_{j \in [N]} \frac{w_i(\sigma_N)}{q_\lambda(x_i)^{1/2}} k(x_i, .) \right\|_{\mathcal{F}}^2 = \mathcal{O}(\sigma_N).$$

Kernel quadrature: ridge leverage scores sampling

$\mathcal{X}$	$\mathcal{F}$ 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.
...	...	...	...

Kernel quadrature: ridge leverage scores sampling

$\mathcal{X}$	$\mathcal{F}$ 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.
...	...	...	...

Challenges

- The theoretical analysis is intricated and requires $\lambda > 0$
- The RLS distribution q_λ is not tractable in general

Kernel quadrature: an alternative analysis using DPPs?

Contributions

- The theoretical analysis for $\lambda = 0$
- Sampling is possible if the spectral decomposition of $\mathbf{\Sigma}$ is known
- Approximate sampling is possible if the spectral decomposition of $\mathbf{\Sigma}$ is **not** tractable

We replace the optimization problem

$$\min_{\boldsymbol{w} \in \mathbb{R}^N} \left\| \mu_g - \sum_{j \in [N]} \frac{w_i}{q(x_i)^{1/2}} k(x_i, .) \right\|_{\mathcal{F}}^2 + \lambda N \|\boldsymbol{w}\|_2^2,$$

by

$$\min_{\boldsymbol{w} \in \mathbb{R}^N} \left\| \mu_g - \sum_{j \in [N]} w_i k(x_i, .) \right\|_{\mathcal{F}}^2.$$

Outline

The determinantal distributions: definition

Definition-Theorem

Let $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ be a kernel s.t. $\int_{\mathcal{X}} \kappa(x,x)\mathrm{d}\omega(x) < +\infty$.
The function

$$f_\kappa(x_1, \dots, x_N) \propto \operatorname{Det} \boldsymbol{\kappa}(\boldsymbol{x})$$

is a p.d.f. on $\mathcal{X}^N$. We denote by Z_κ its normalization constant.

We study two cases:

- Projection DPP:

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

- Continuous volume sampling (CVS):

$$\kappa(x,y) = k(x,y) = \sum_{m \in \mathbb{N}^*} \sigma_m e_m(x) e_m(y)$$

Main results: the case of the projection DPP

The theoretical guarantee in the case $\kappa = \mathfrak{K}$ is given in the following result.

Theorem (B., Bardenet and Chainais (2019))

Define $r_N = \sum_{m \geq N+1} \sigma_m$. *Then*

$$\mathbb{E}_{\mathfrak{K}} \sup_{\|g\|_{\omega} \leq 1} \|\mu_g - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})}\mu_g\|_{\mathcal{F}}^2 \leq 4N^2 r_N.$$

Examples:

σ_N	$N^2 r_N$	Empirical rate
N^{-2s}	$N^3 \mathcal{O}(\sigma_{N+1})$	$\mathcal{O}(\sigma_{N+1})$
α^N	$N^2 \mathcal{O}(\sigma_{N+1})$ $\approx \mathcal{O}(\sigma_{N+1})$	$\mathcal{O}(\sigma_{N+1})$

Main results: a lower bound

Theorem (Pinkus (1985))

Assume that $\boldsymbol{\Sigma}$ is compact, then

$$\inf_{\substack{\mathcal{Y} \subset \mathcal{F} \\ \dim \mathcal{Y} = N}} \sup_{\|g\|_\omega \leq 1} \|\mu_g - \boldsymbol{\Pi}_{\mathcal{Y}} \mu_g\|_{\mathcal{F}}^2 = \sigma_{N+1}$$

Corollary

For any configuration $\boldsymbol{x} \in \mathcal{X}^N$ such that $\dim \mathcal{T}(\boldsymbol{x}) = N$,

$$\sup_{\|g\|_\omega \leq 1} \|\mu_g - \boldsymbol{\Pi}_{\mathcal{T}(\boldsymbol{x})} \mu_g\|_{\mathcal{F}}^2 \geq \sigma_{N+1}$$

Main results: a tractable formula under volume sampling

The theoretical guarantee in the case $\kappa = k$ is given in the following result.

Theorem (B., Bardenet and Chainais (2020))

Let $g = \sum_{m \in \mathbb{N}^*} \langle g, e_m \rangle_\omega e_m$ *then*

$$\mathbb{E}_k \|\mu_g - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})} \mu_g\|_{\mathcal{F}}^2 = \sum_{m \in \mathbb{N}^*} \langle g, e_m \rangle_\omega^2 \epsilon_m(N),$$

$$\epsilon_m(N) = \mathbb{E}_k \|\mu_{e_m} - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})} \mu_{e_m}\|_{\mathcal{F}}^2 = \sigma_m \frac{\sum_{|T|=N, m \notin T} \prod_{t \in T} \sigma_t}{\sum_{|T|=N} \prod_{t \in T} \sigma_t}.$$

How good is it?

Main results: how large are the epsilons?

Theorem (B., Bardenet and Chainais (2020))

If there exists $B > 0$ such that $\min_{M \in [N]} \frac{\sum_{m \geq M} \sigma_m}{(N-M+1)\sigma_{N+1}} \leq B.$

Then

$$\sup_{\|g\|_\omega \leq 1} \mathbb{E}_k \|\mu_g - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})}\mu_g\|_{\mathcal{F}}^2 = \epsilon_1(N) \leq (1+B)\sigma_{N+1}.$$

Examples:

σ_N	B
N^{-2s}	$(1+1/(2s-1))^{2s}$
α^N	$\alpha/(1-\alpha)$

Main results: how large are the epsilons?

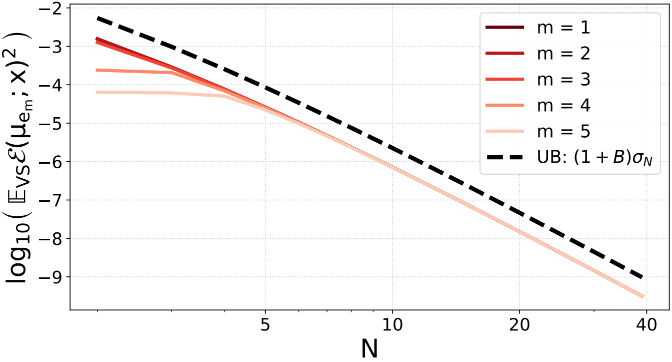


Figure: $\log_{10} \epsilon_m(N)$ as a function of N when $\sigma_N = N^{-2s}$, with $s = 3$.

Main results: how large are the epsilons?

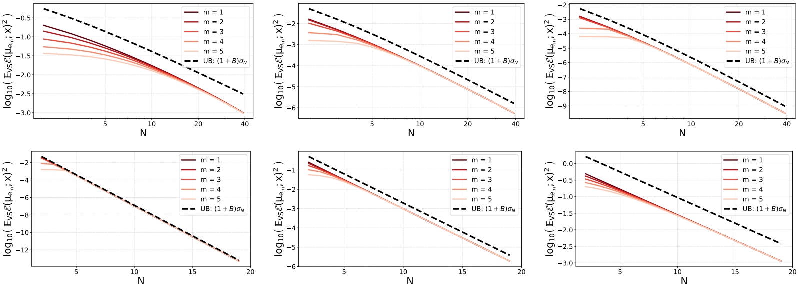


Figure: Other examples in different RKHSs.

Main results: a summary

Quadrature	Distribution	Theoretical rate	Empirical rate
EZQ	DPP	$\mathcal{O}(r_{N+1})$ [B. (2021)]	$\mathcal{O}(r_{N+1})$ [B. (2021)]
OKQ	DPP	$N^2\mathcal{O}(r_{N+1})$ [B. et al. (2019)] $\mathcal{O}(r_{N+1})$ [B. (2021)]	$\mathcal{O}(\sigma_{N+1})$ [B. et al. (2019)]
OKQ	CVS	$\mathcal{O}(\sigma_{N+1})$ [B. et al. (2020)]	$\mathcal{O}(\sigma_{N+1})$ [B. et al. (2020)]

Main results: interpolation beyond quadrature

For $f \in \mathcal{F}$, we have

$$f = \sum_{m \in \mathbb{N}^*} \sqrt{\sigma_m} \langle f, e_m^{\mathcal{F}} \rangle_{\mathcal{F}} e_m = \mathbf{\Sigma}^{1/2} \tilde{f}$$

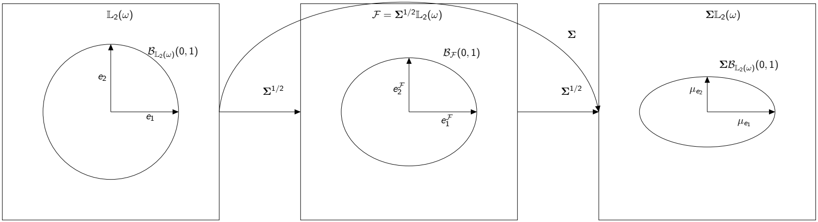
with $\tilde{f} := \sum_{m \in \mathbb{N}^*} \langle f, e_m^{\mathcal{F}} \rangle_{\mathcal{F}} e_m \in \mathbb{L}_2(\omega)$.

Main results: interpolation beyond quadrature

For $f \in \mathcal{F}$, we have

$$f = \sum_{m \in \mathbb{N}^*} \sqrt{\sigma_m} \langle f, e_m^{\mathcal{F}} \rangle_{\mathcal{F}} e_m = \boldsymbol{\Sigma}^{1/2} \tilde{f}$$

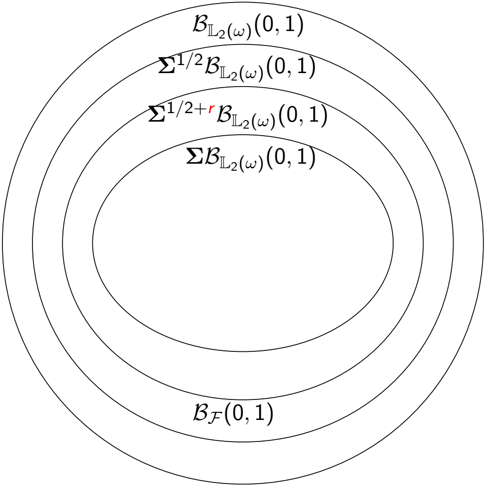
with $\tilde{f} := \sum_{m \in \mathbb{N}^*} \langle f, e_m^{\mathcal{F}} \rangle_{\mathcal{F}} e_m \in \mathbb{L}_2(\omega)$.



The embeddings μ_g belongs to an ellipsoid in $\mathbb{L}_2(\omega)$

Main results: interpolation beyond quadrature

We can extend the previous result outside $\boldsymbol{\Sigma}\mathbb{L}_2(\omega)$.



We prove the rate $\mathcal{O}(\sigma_{N+1}^{2r})$ in $\boldsymbol{\Sigma}^{1/2+r}\mathbb{L}_2(\omega)$ for $r \in [0, 1/2]$.

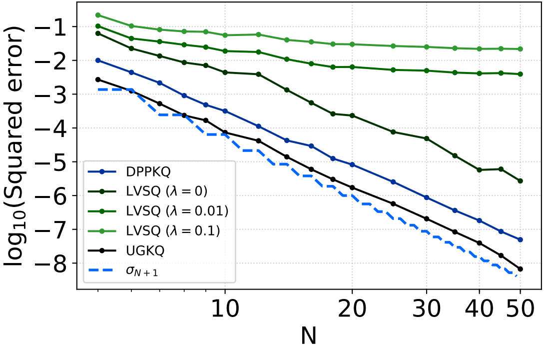
Outline

Numerical simulations: DPP in the periodic Sobolev space

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

$$\mathbb{E}_{\kappa} \sup_{\|g\|_{\omega} \leq 1} \|\mu_g - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})}\mu_g\|_{\mathcal{F}}^2 \approx \mathbb{E}_{\kappa} \sup_{g \in \mathcal{G}} \|\mu_g - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})}\mu_g\|_{\mathcal{F}}^2$$

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 $s = 3$.



We report $\epsilon_m(N) = \mathbb{E}_\kappa \|\mu_{e_m} - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})}\mu_{e_m}\|^2_{\mathcal{F}}$, where $\mathcal{F}$ is the $\otimes$ of Sobolev spaces (the Korobov space) of dimension $d = 2$ and order $s = 1$.

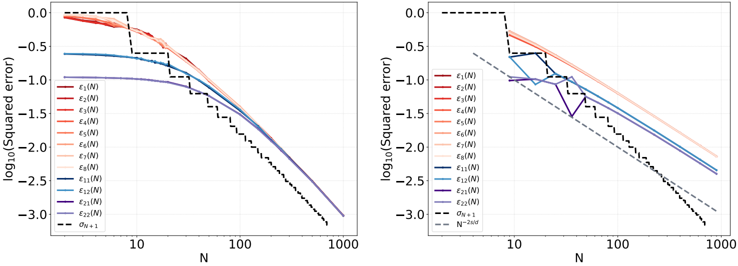


Figure: DPPKQ (left) vs OKQ on the uniform grid (right)

Numerical simulations: the Gaussian space

We report the interpolation error for $g \in \{e_1, e_{15}\}$, $\mathcal{F}$ is the Gaussian space corresponding to the Gaussian kernel and ω is the Gaussian measure.

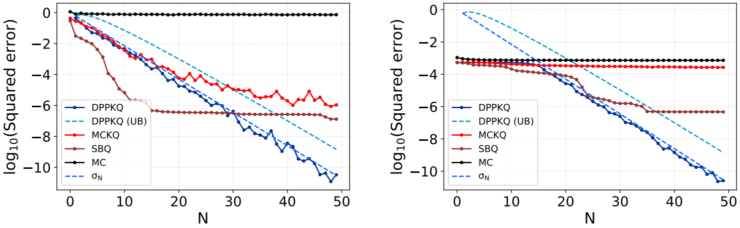
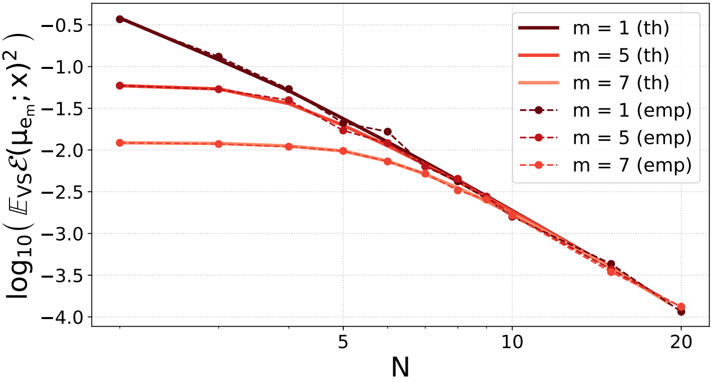


Figure: The squared interpolation error for e_1 (Left), vs e_{15} (Right).

Numerical simulations: CVS in the periodic Sobolev space

We report the empirical expectation of the square of the interpolation error $\mathbb{E}_{\kappa}\|\mu_g - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})}\mu_g\|_{\mathcal{F}}^2$ for CVS ($\kappa = k$) in the periodic Sobolev space of order $s = 2$ and $g \in \{e_1, e_5, e_7\}$.



Outline

Intuitions

Observe that

$$\begin{aligned}\mathbb{E}_\kappa\|\mu_g - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})}\mu_g\|_{\mathcal{F}}^2 &= \mathbb{E}_\kappa\|\boldsymbol{O}_{\boldsymbol{x}}\mathbf{\Sigma} g\|_{\mathcal{F}}^2 \\ &= \mathbb{E}_\kappa\|\boldsymbol{O}_{\boldsymbol{x}}\mathbf{\Sigma}_N g + \boldsymbol{O}_{\boldsymbol{x}}\mathbf{\Sigma}_N^\perp g\|_{\mathcal{F}}^2 \\ &\leq 2\Big(\mathbb{E}_\kappa\|\boldsymbol{O}_{\boldsymbol{x}}\mathbf{\Sigma}_N g\|_{\mathcal{F}}^2 + \mathbb{E}_\kappa\|\boldsymbol{O}_{\boldsymbol{x}}\mathbf{\Sigma}_N^\perp g\|_{\mathcal{F}}^2\Big)\end{aligned}$$

where

$$\boldsymbol{O}_{\boldsymbol{x}} = \mathbb{I}_{\mathcal{F}} - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})} = \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})^\perp},$$

$$\mathbf{\Sigma}_N = \sum_{m=1}^{N}\sigma_m e_m \otimes e_m, \qquad \mathbf{\Sigma}_N^\perp = \sum_{m=N+1}^{+\infty}\sigma_m e_m \otimes e_m.$$

Intuitions

Observe that

$$\begin{aligned}\mathbb{E}_\kappa\|\mu_g - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})}\mu_g\|_{\mathcal{F}}^2 &= \mathbb{E}_\kappa\|\boldsymbol{O}_{\boldsymbol{x}}\mathbf{\Sigma} g\|_{\mathcal{F}}^2\\ &= \mathbb{E}_\kappa\|\boldsymbol{O}_{\boldsymbol{x}}\mathbf{\Sigma}_N g + \boldsymbol{O}_{\boldsymbol{x}}\mathbf{\Sigma}_N^\perp g\|_{\mathcal{F}}^2\\ &\leq 2\Big(\mathbb{E}_\kappa\|\boldsymbol{O}_{\boldsymbol{x}}\mathbf{\Sigma}_N g\|_{\mathcal{F}}^2 + \mathbb{E}_\kappa\|\boldsymbol{O}_{\boldsymbol{x}}\mathbf{\Sigma}_N^\perp g\|_{\mathcal{F}}^2\Big)\end{aligned}$$

where

$$\boldsymbol{O}_{\boldsymbol{x}} = \mathbb{I}_{\mathcal{F}} - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})} = \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})^\perp},$$

$$\mathbf{\Sigma}_N = \sum_{m=1}^{N}\sigma_m e_m\otimes e_m, \qquad \mathbf{\Sigma}_N^\perp = \sum_{m=N+1}^{+\infty}\sigma_m e_m\otimes e_m.$$

$\boldsymbol{O}_{\boldsymbol{x}} = \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})^\perp}$ is an orthogonal projection, then

$$\|\boldsymbol{O}_{\boldsymbol{x}}\mathbf{\Sigma}_N^\perp g\|_{\mathcal{F}}^2 \leq \|\mathbf{\Sigma}_N^\perp g\|_{\mathcal{F}}^2 = \sum_{m\geq N+1}\sigma_m\langle g, e_m\rangle_\omega^2 \leq \sigma_{N+1}\|g\|_\omega^2.$$

Intuitions

Let $m \in \mathbb{N}^*$ and put $g = e_m$

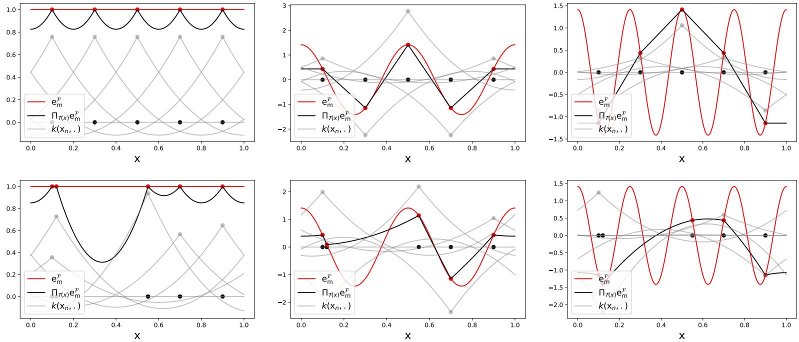
$$\|\boldsymbol{O}_{\boldsymbol{x}} \boldsymbol{\Sigma}_N e_m\|_{\mathcal{F}}^2 = \sigma_m \|\boldsymbol{O}_{\boldsymbol{x}} e_m^{\mathcal{F}}\|_{\mathcal{F}}^2 = \sigma_m \|e_m^{\mathcal{F}} - \boldsymbol{\Pi}_{\mathcal{T}(\boldsymbol{x})} e_m^{\mathcal{F}}\|_{\mathcal{F}}^2$$

The error term is the product of two terms:

- the eigenvalue σ_m
- the reconstruction term $\|e_m^{\mathcal{F}} - \boldsymbol{\Pi}_{\mathcal{T}(\boldsymbol{x})} e_m^{\mathcal{F}}\|_{\mathcal{F}}^2 \in [0, 1]$

Intuitions

$$\sigma_m \| e_m^{\mathcal{F}} - \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})} e_m^{\mathcal{F}} \|_{\mathcal{F}}^2 = \sigma (1 - \| \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})} e_m^{\mathcal{F}} \|_{\mathcal{F}}^2)$$



Intuitions

Theorem

Under the distribution of CVS $\kappa = k$, we have

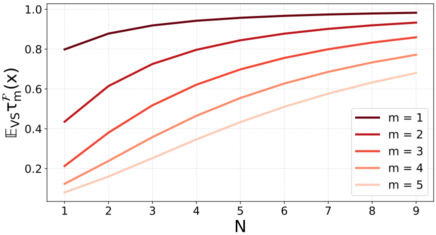
$$\forall m \in \mathbb{N}^*, \;\; \mathbb{E}_k \|\mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})} e_m^{\mathcal{F}}\|_{\mathcal{F}}^2 = \frac{\sum\limits_{|T|=N, m \in T} \prod\limits_{t \in T} \sigma_t}{\sum\limits_{|T|=N} \prod\limits_{t \in T} \sigma_t},$$

and

$$\forall m \neq m' \in \mathbb{N}^*, \;\; \mathbb{E}_k \langle \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})} e_m^{\mathcal{F}}, \mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})} e_{m'}^{\mathcal{F}} \rangle_{\mathcal{F}} = 0.$$

Intuitions

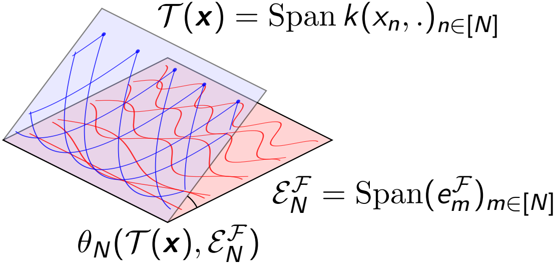
$$\mathbb{E}_k \tau_m^{\mathcal{F}}(\boldsymbol{x}) := \mathbb{E}_k \|\mathbf{\Pi}_{\mathcal{T}(\boldsymbol{x})} e_m^{\mathcal{F}}\|_{\mathcal{F}}^2 = \frac{\sum\limits_{|T|=N, m\in T} \prod\limits_{t\in T} \sigma_t}{\sum\limits_{|T|=N} \prod\limits_{t\in T} \sigma_t}.$$



Under CVS, $\mathcal{T}(\boldsymbol{x})$ gets closer to $\mathcal{E}_N = \mathrm{Span}(e_m^{\mathcal{F}})_{m\in[N]}$ as $N \to +\infty$

Intuitions

Alternatively, we can quantify the proximity between the subspaces $\mathcal{T}(\boldsymbol{x})$ and $\mathcal{E}_N^{\mathcal{F}}$ using **the principal angles** $(\theta_i(\mathcal{T}(\boldsymbol{x}), \mathcal{E}_N^{\mathcal{F}}))_{i \in [N]}$.



For example, we have

$$\sup_{m \in [N]} \|e_m^{\mathcal{F}} - \boldsymbol{\Pi}_{\mathcal{T}(\boldsymbol{x})} e_m^{\mathcal{F}}\|_{\mathcal{F}}^2 \leq \frac{1}{\cos^2 \theta_N(\mathcal{T}(\boldsymbol{x}), \mathcal{E}_N^{\mathcal{F}})} - 1.$$

Intuitions

Theorem (B., Bardenet and Chainais (2019))

For $N \in \mathbb{N}^*$

$$\mathbb{E}_{\mathfrak{K}} \prod_{n=1}^{N} \frac{1}{\cos^2 \theta_n\Big(\mathcal{E}_N^{\mathcal{F}}, \mathcal{T}(\boldsymbol{x})\Big)} = \frac{1}{\prod\limits_{n=1}^{N} \sigma_n} \sum_{\substack{T \subset \mathbb{N}^* \\ |T|=N}} \prod_{t \in T} \sigma_t,$$

and

$$\mathbb{E}_{\mathfrak{K}} \sum_{n=1}^{N} \frac{1}{\cos^2 \theta_n\Big(\mathcal{E}_N^{\mathcal{F}}, \mathcal{T}(\boldsymbol{x})\Big)} = N + \sum_{v \in [N]} \frac{1}{\sigma_v} \sum_{w \in \mathbb{N}^* \smallsetminus [N]} \sigma_w.$$

Outline

The determinantal distributions: sequential sampling

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

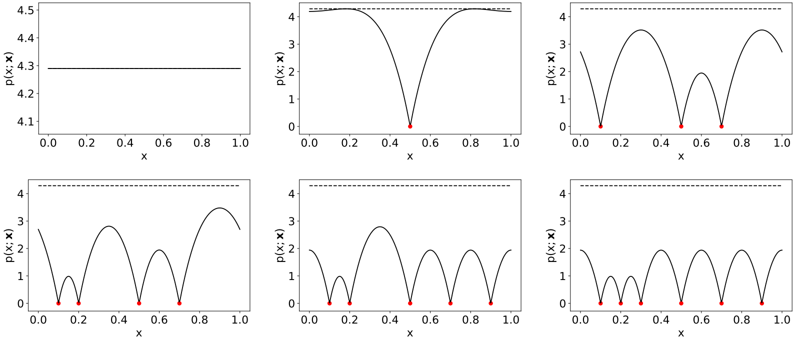
$$\begin{aligned}
\mathrm{Det}\, \boldsymbol{\kappa}(\boldsymbol{x}) = & \kappa(x_1, x_1) \\
& \times \left(\kappa(x_2, x_2) - \frac{\kappa(x_1, x_2)^2}{\kappa(x_1, x_1)} \right) \\
& \cdots \\
& \times \left(\kappa(x_\ell, x_\ell) - \phi_{\boldsymbol{x}_\ell}(x_\ell)^T \boldsymbol{\kappa}(\boldsymbol{x}_\ell)^{-1}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

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

The determinantal distributions: sequential sampling

$$p(x;\boldsymbol{x}) = \kappa(x,x) - \phi_{\boldsymbol{x}}(x)^\intercal \boldsymbol{\kappa}(\boldsymbol{x})^{-1}\phi_{\boldsymbol{x}}(x),$$

$$\phi_{\boldsymbol{x}}(x) = (k(x,\xi))_{\xi\in\boldsymbol{x}},$$



The determinantal distributions: sequential sampling

If κ is a projection kernel

$$\int_{\mathcal{X}} p(x; \boldsymbol{x}) \mathrm{d}\omega(x) = N - |\boldsymbol{x}|,$$

and

$$f_\kappa(\boldsymbol{x}) = \frac{1}{N!} \operatorname{Det} \boldsymbol{\kappa}(\boldsymbol{x}) = \prod_{\ell \in [N]} \frac{1}{N - \ell + 1} p(x_\ell; \boldsymbol{x}_\ell)$$

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

The determinantal distributions: sampling

If $\kappa = k$, the sequential algorithm is an approximation

Theorem (Rezaei and Gharan (2019))

Let $\boldsymbol{x}$ the output of the sequential algorithm for $\kappa = k$, then $\boldsymbol{x}$ follows the density f_{seq} that satisfies

$$f_{\mathrm{seq}}(\boldsymbol{x}) \leq N!^2 f_k(\boldsymbol{x}).$$

An MCMC algorithm for CVS [Rezaei and Gharan (2019)]

CVS is the stationary distribution of a Markov chain that can be implemented in a **fully kernelized** way: using only the evaluations of the kernel k. f_{seq} is the initialization of the Markov Chain.

Outline

Conclusion

Take-home messages

- The theoretical analysis for $\lambda = 0$
- Sampling is possible if the spectral decomposition of $\mathbf{\Sigma}$ is known
- Approximate sampling is possible if the spectral decomposition of $\mathbf{\Sigma}$ is **not** tractable

Perspectives

- Efficient sampling projection DPPs and/or CVS?
- Quadratures on manifolds?
- Extension to random features?
- The theoretical analysis of the stability

Perspectives

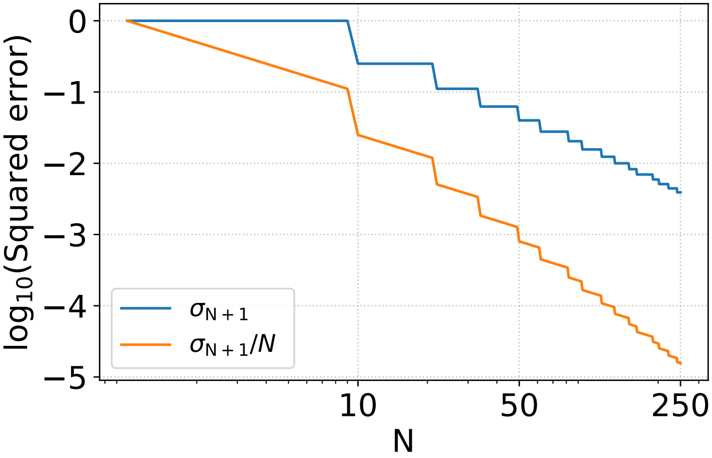
The quadrature	OKQ	The linear statistic
	This work	Bardenet and Hardy (2016)
The expression	$\sum_{i \in [N]} \hat{w}_i(\boldsymbol{x}) f(x_i)$	$\sum_{i \in [N]} f(x_i)/\kappa(x_i, x_i)$
$\mathcal{F}$	RKHS	Not an RKHS
		(Sobolev spaces with $s \leq \frac{d}{2}$)
$\|.\|_{\mathcal{F}}$	$\sum_{m \in \mathbb{N}^*} \sigma_m < +\infty$	$\sum_{m \in \mathbb{N}^*} \sigma_m = +\infty$
Convergence rate	σ_{N+1}	σ_{N+1}/N
Non asymptotic	Yes	No
g	$\in \mathbb{L}_2(\omega)$	$\equiv 1$

A universal construction of quadrature rules that achieve the rate σ_{N+1}/N on RKHSs?

$$\sigma_{N+1} \times 1/N$$

Smoothness in $\mathcal{F}$ Monte Carlo rate

Example: $\mathcal{F}$ is the Korobov space of dimension $d = 2$ ($s = 1$).



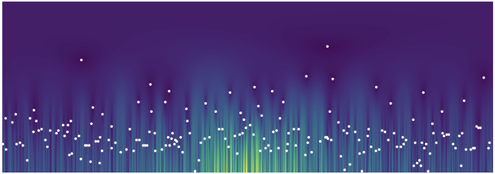
A workshop about DPPs, quantum physics, and signal processing at ENS de Lyon in two weeks!

Determinantal and permanental point processes, quantum physics, and signal processing (dpp-fermions-2022)

30 May 2022 to 10 June 2022
ENS de Lyon
Europe/Paris timezone

Enter your search term

- Overview
- Scientific Programme
- Timetable
- Registration
- Practical information



Since their introduction by Odile Macchi in 1975 to model the spatial distribution of fermions and bosons in optical beams, determinantal and permanental point processes have received a lot of attention due to their connections with random matrices, statistical physics and more recently signal processing and machine learning, where the repulsive properties of DPP allow for instance to improve the efficiency of Monte Carlo methods. Quantum physics and signal processing are also deeply related through time-frequency analysis, which provides tools to understand complex optical systems.

Thank you for your attention!