Quadrature in reproducing kernel Hilbert spaces with repulsive point processes

Rémi Bardenet

CNRS & CRIStAL, Univ. Lille, France

Joint work with

Figure: Adrien Hardy, Ayoub Belhadji, Pierre Chainais

Prologue: Determinantal point processes

Numerical integration

Tight interpolation rates in RKHSs

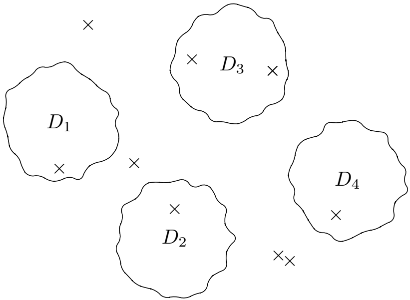
Prologue: Determinantal point processes

Numerical integration

Tight interpolation rates in RKHSs

How do you describe a point process?

- Let $\mathcal{X}$ be a complete metric space, with μ a Borel measure.
- A point process is a random configuration of points in $\mathcal{X}$.

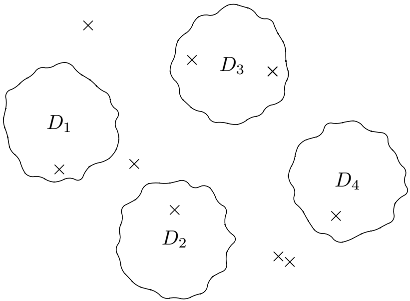


- Correlation intensitities (ρ_n), are defined by

$$\mathbb{E}\left[N(D_1)\dots N(D_n)\right] = \int \rho_n(x_1,\dots,x_n)\mathrm{d}\mu(x_1)\dots\mathrm{d}\mu(x_n), \quad n \geqslant 1.$$

How do you describe a point process?

- Let $\mathcal{X}$ be a complete metric space, with μ a Borel measure.
- A point process is a random configuration of points in $\mathcal{X}$.



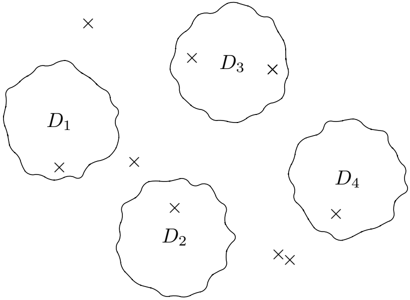
- Correlation intensitities (ρ_n), are defined by

$$\mathbb{E}\left[N(D_1)\dots N(D_n)\right] = \int \rho_n(x_1,\dots,x_n)\mathrm{d}\mu(x_1)\dots\mathrm{d}\mu(x_n), \quad n \geqslant 1.$$

- The Poisson process corresponds, e.g., to $\rho_n(x_1,\dots,x_n) = \lambda(x_1)\dots\lambda(x_n)$, $\mathrm{d}\mu = \mathrm{d}x$.

How do you describe a point process?

- Let $\mathcal{X}$ be a complete metric space, with μ a Borel measure.
- A point process is a random configuration of points in $\mathcal{X}$.



- Correlation intensitities (ρ_n), are defined by

$$\mathbb{E}\left[N(D_1)\dots N(D_n)\right] = \int \rho_n(x_1,\dots,x_n)\mathrm{d}\mu(x_1)\dots\mathrm{d}\mu(x_n), \quad n \geqslant 1.$$

- A DPP is defined by $\rho_n(x_1,\dots,x_n) = \det\left(K(x_i,x_j)\right)_{1\leqslant i,j\leqslant n}$

Basic properties, assuming existence

- $\mathbb{E}[N(X)] = \int K(x,x)\mathrm{d}\mu(x)$.

- If $K(x,y)$ is the kernel of a projection of rank r in $L^2(\mu)$, then $X \sim \mathrm{DPP}(K,\mu)$ has cardinality r, almost surely.

- Interaction can be read in

$$\begin{aligned} \rho_2(x,y) &= K(x,x)K(y,y) - K(x,y)K(y,x) \\ &= \rho_1(x)\rho_1(y) - |K(x,y)|^2 \leqslant \rho_1(x)\rho_1(y). \end{aligned}$$

  if $K(x,y) = \overline{K(y,x)}$.

- For finite $\mathcal{X}$, take μ to be the counting measure, the correlation intensities read

$$\mathbb{P}_{X\sim\mathrm{DPP}(K,\mu)}(A \subset X) = \det K(i,j)_{i,j\in A} = \det \mathbf{K}_A.$$

- $\mathbb{E}[N(X)] = \int K(x,x)\mathrm{d}\mu(x)$.

- If $K(x,y)$ is the kernel of a projection of rank r in $L^2(\mu)$, then $X \sim \mathrm{DPP}(K,\mu)$ has cardinality r, almost surely.

- Interaction can be read in

$$\begin{aligned}\rho_2(x,y) &= K(x,x)K(y,y) - K(x,y)K(y,x)\\ &= \rho_1(x)\rho_1(y) - |K(x,y)|^2 \leqslant \rho_1(x)\rho_1(y).\end{aligned}$$

  if $K(x,y) = \overline{K(y,x)}$.

- For finite $\mathcal{X}$, take μ to be the counting measure, the correlation intensities read

$$\mathbb{P}_{X\sim \mathrm{DPP}(K,\mu)}(A\subset X) = \det K(i,j)_{i,j\in A} = \det \mathbf{K}_A.$$

Basic properties, assuming existence

- $\mathbb{E}[N(X)] = \int K(x,x)\mathrm{d}\mu(x)$.
- If $K(x,y)$ is the kernel of a projection of rank r in $L^2(\mu)$, then $X \sim \mathrm{DPP}(K,\mu)$ has cardinality r, almost surely.
- Interaction can be read in

$$
\begin{aligned}
\rho_2(x,y) &= K(x,x)K(y,y) - K(x,y)K(y,x) \\
&= \rho_1(x)\rho_1(y) - |K(x,y)|^2 \leqslant \rho_1(x)\rho_1(y).
\end{aligned}
$$

if $K(x,y) = \overline{K(y,x)}$.

- For finite $\mathcal{X}$, take μ to be the counting measure, the correlation intensities read

$$\mathbb{P}_{X\sim\mathrm{DPP}(K,\mu)}(A \subset X) = \det K(i,j)_{i,j\in A} = \det \mathbf{K}_A.$$

- $\mathbb{E}[N(X)] = \int K(x,x)\mathrm{d}\mu(x)$.
- If $K(x,y)$ is the kernel of a projection of rank r in $L^2(\mu)$, then $X \sim \mathrm{DPP}(K,\mu)$ has cardinality r, almost surely.
- Interaction can be read in

$$\begin{aligned}\rho_2(x,y) &= K(x,x)K(y,y) - K(x,y)K(y,x) \\ &= \rho_1(x)\rho_1(y) - |K(x,y)|^2 \leqslant \rho_1(x)\rho_1(y).\end{aligned}$$

if $K(x,y) = \overline{K(y,x)}$.

- For finite $\mathcal{X}$, take μ to be the counting measure, the correlation intensities read

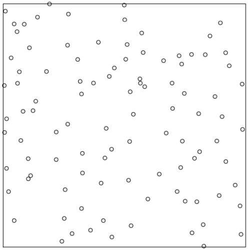
$$\mathbb{P}_{X\sim\mathrm{DPP}(K,\mu)}(A \subset X) = \det K(i,j)_{i,j\in A} = \det \mathbf{K}_A.$$

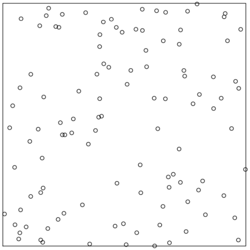
Theorem (Macchi 1975; Soshnikov 2002)

If K defines a self-adjoint, trace-class operator on $L^2(\mu)$, then $DPP(K, \mu)$ exists iff the spectrum of K in is $[0, 1]$.

- For instance, take $K(x, y) = \rho \exp(-\|x - y\|^2/\alpha^2)$ and μ Lebesgue.
- Fourier arguments[1] show that the DPP exists iff

$$\rho(\sqrt{\pi}\alpha)^d \exp(-\|\pi\alpha x\|^2) \leqslant 1,$$





[1] Lavancier, Møller, and Rubak 2014.

A constructive proof of Macchi-Soshnikov[2]

Let $K(x,y) = \sum_k \lambda_k \varphi_k(x)\varphi_k(y)$.

1. sample $B_k \sim \mathrm{Ber}(\lambda_k)$, for all k.
2. letting $N = \sum_k B_k$ and

$$\tilde{K}(x,y) = \sum_{k \geqslant 0} B_k \varphi_k(x)\varphi_k(y),$$

   sample

$$x_1, \dots, x_N \sim \det\left(\tilde{K}(x_i, x_j)\right)_{1 \leqslant i,j \leqslant N}$$

[2] Hough, Krishnapur, Peres, and Virág 2006.

Let $K(x,y) = \sum_k \lambda_k \varphi_k(x)\varphi_k(y)$.

1. sample $B_k \sim \mathrm{Ber}(\lambda_k)$, for all k.
2. letting $N = \sum_k B_k$ and

$$\tilde{K}(x,y) = \sum_{k \geqslant 0} B_k \varphi_k(x)\varphi_k(y),$$

   sample

$$x_1, \dots, x_N \sim \det\left(\tilde{K}(x_i, x_j)\right)_{1 \leqslant i,j \leqslant N}$$

[2] Hough, Krishnapur, Peres, and Virág 2006.

Let $K(x,y) = \sum_k \lambda_k \varphi_k(x)\varphi_k(y)$.

1. sample $B_k \sim \mathrm{Ber}(\lambda_k)$, for all k.
2. letting $N = \sum_k B_k$ and

$$\tilde{K}(x,y) = \sum_{k \geqslant 0} B_k \varphi_k(x)\varphi_k(y),$$

   sample

$$x_1, \dots, x_N \sim \det\left(\tilde{K}(x_i, x_j)\right)_{1 \leqslant i,j \leqslant N}$$

[2] Hough, Krishnapur, Peres, and Virág 2006.

A constructive proof of Macchi-Soshnikov[2]

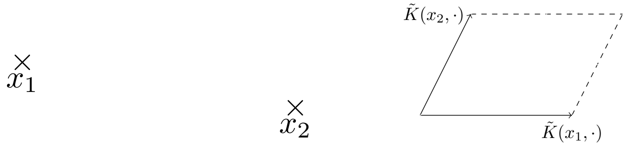
Let $K(x, y) = \sum_k \lambda_k \varphi_k(x)\varphi_k(y)$.

1. sample $B_k \sim \mathrm{Ber}(\lambda_k)$, for all k.
2. letting $N = \sum_k B_k$ and

$$\tilde{K}(x, y) = \sum_{k \geqslant 0} B_k \varphi_k(x)\varphi_k(y),$$

   sample

$$x_1, \dots, x_N \sim \det\left(\tilde{K}(x_i, x_j)\right)_{1 \leqslant i,j \leqslant N}$$



[2] Hough, Krishnapur, Peres, and Virág 2006.

A constructive proof of Macchi-Soshnikov[2]

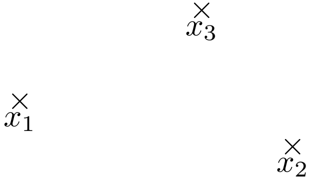
Let $K(x,y) = \sum_k \lambda_k \varphi_k(x)\varphi_k(y)$.

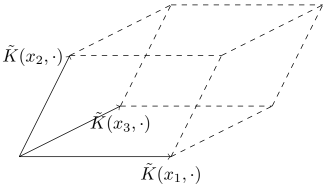
1. sample $B_k \sim \mathrm{Ber}(\lambda_k)$, for all k.
2. letting $N = \sum_k B_k$ and

$$\tilde{K}(x,y) = \sum_{k \geqslant 0} B_k \varphi_k(x)\varphi_k(y),$$

   sample

$$x_1, \ldots, x_N \sim \det\left(\tilde{K}(x_i, x_j)\right)_{1 \leqslant i,j \leqslant N}$$





[2] Hough, Krishnapur, Peres, and Virág 2006.

Prologue: Determinantal point processes

Numerical integration

Tight interpolation rates in RKHSs

The goal is to approximate

$$\int f \,\mathrm{d}\mu = \int f(x)\omega(x)\mathrm{d}x \approx \sum_{i=1}^{N} w_i f(x_i).$$

- How to choose the nodes x_i?
- How to choose the weights w_i?

Monte Carlo integration (importance sampling, MCMC, etc.)

- Choose the nodes randomly, and the weights $w_i = w(x_i, x_{-i})$.
- Typical error is

$$\sqrt{\mathbb{E}\left[\int f \,\mathrm{d}\mu - \sum_{i=1}^{N} w_i f(x_i)\right]^2} \sim \frac{1}{\sqrt{N}}.$$

Projection DPPs

- Let $(\varphi_k)_{k=0,\dots,N-1}$ be an orthonormal sequence in $L^2(\mu)$.
- Let $\mathrm{K}(x,y) = \sum_{k=0}^{N-1} \varphi_k(x)\varphi_k(y)$.

Definition (Hough, Krishnapur, Peres, and Virág 2006)

$X = \{x_1, \dots, x_N\}$ is the DPP with kernel K and reference measure μ if

$$x_1, \dots, x_N \sim \frac{1}{N!} \det \left[\mathrm{K}(x_i, x_\ell)\right]_{i,\ell=1}^{N} \mathrm{d}\mu(x_1) \dots \mathrm{d}\mu(x_N).$$

1. If $\mu = \sum_{x \in \mathcal{X}} \delta_x$, one recovers

$$\mathbb{P}(A \subset X) = \det \mathbf{K}_A.$$

2. $x_1 \sim \frac{1}{N}\mathrm{K}(x,x)\mathrm{d}\mu(x)$ so that $\mathbb{E}\sum_{i=1}^N \frac{f(x_i)}{\mathrm{K}(x_i,x_i)} = \int f \,\mathrm{d}\mu$.
3. A natural choice of $\varphi_k : \mathbb{R}^d \to \mathbb{R}$ is orthogonal polynomials w.r.t. μ.

Projection DPPs

- Let $(\varphi_k)_{k=0,\dots,N-1}$ be an orthonormal sequence in $L^2(\mu)$.
- Let $\mathrm{K}(x,y) = \sum_{k=0}^{N-1} \varphi_k(x)\varphi_k(y)$.

Definition (Hough, Krishnapur, Peres, and Virág 2006)

$X = \{x_1, \dots, x_N\}$ is the DPP with kernel K and reference measure μ if

$$x_1, \dots, x_N \sim \frac{1}{N!} \det \left[\mathrm{K}(x_i, x_\ell)\right]_{i,\ell=1}^{N} \mathrm{d}\mu(x_1) \dots \mathrm{d}\mu(x_N).$$

1. If $\mu = \sum_{x\in\mathcal{X}} \delta_x$, one recovers

$$\mathbb{P}(A \subset X) = \det \mathbf{K}_A.$$

2. $x_1 \sim \frac{1}{N}\mathrm{K}(x,x)\mathrm{d}\mu(x)$ so that $\mathbb{E}\sum_{i=1}^{N} \frac{f(x_i)}{\mathrm{K}(x_i,x_i)} = \int f \mathrm{d}\mu$.
3. A natural choice of $\varphi_k : \mathbb{R}^d \to \mathbb{R}$ is orthogonal polynomials w.r.t. μ.

- Let $(\varphi_k)_{k=0,\dots,N-1}$ be an orthonormal sequence in $L^2(\mu)$.
- Let $\mathrm{K}(x,y) = \sum_{k=0}^{N-1} \varphi_k(x)\varphi_k(y)$.

Definition (Hough, Krishnapur, Peres, and Virág 2006)

$X = \{x_1, \dots, x_N\}$ is the DPP with kernel K and reference measure μ if

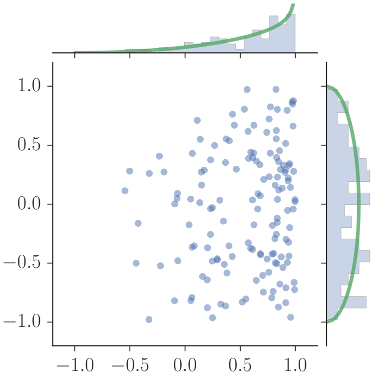
$$x_1, \dots, x_N \sim \frac{1}{N!} \det \left[\mathrm{K}(x_i, x_\ell)\right]_{i,\ell=1}^{N} \mathrm{d}\mu(x_1)\dots\mathrm{d}\mu(x_N).$$

1. If $\mu = \sum_{x\in\mathcal{X}} \delta_x$, one recovers

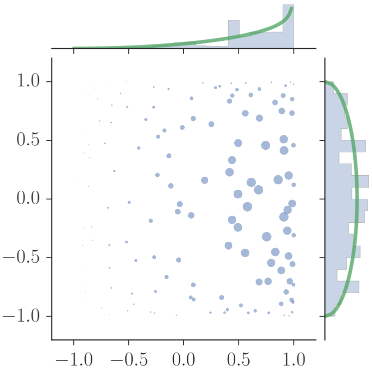
$$\mathbb{P}(A \subset X) = \det \mathbf{K}_A.$$

2. $x_1 \sim \frac{1}{N}\mathrm{K}(x,x)\mathrm{d}\mu(x)$ so that $\mathbb{E}\sum_{i=1}^{N} \frac{f(x_i)}{\mathrm{K}(x_i,x_i)} = \int f\mathrm{d}\mu$.
3. A natural choice of $\varphi_k : \mathbb{R}^d \to \mathbb{R}$ is orthogonal polynomials w.r.t. μ.

Multivariate orthogonal polynomial ensembles[3]



(a) i.i.d.



(b) DPP

[3] Gautier, Bardenet, Polito, and Valko 2019.

Our first Monte Carlo result

Theorem (Bardenet and Hardy 2020)

Let $\mu(\mathrm{d}x) = \omega(x)\mathrm{d}x$ with ω separable, $\mathscr{C}^1$, positive on $(-1,1)^d$, and satisfying a regularity assumption. Let $\varepsilon > 0$. If $x_1, \dots, x_N$ stands for the associated OPE, then for f $\mathscr{C}^1$ vanishing outside $[-1+\varepsilon, 1-\varepsilon]^d$,

$$\sqrt{N^{1+1/d}} \left(\sum_{i=1}^{N} \frac{f(x_i)}{\mathrm{K}(x_i, x_i)} - \int f(x)\mu(\mathrm{d}x) \right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0, \Omega^2_{f,\omega}),$$

where

$$\Omega^2_{f,\omega} = \frac{1}{2} \sum_{k_1,\dots,k_d=0}^{\infty} (k_1 + \cdots + k_d) \widehat{\left(\frac{f\omega}{\omega_{eq}^{\otimes d}}\right)}(k_1, \dots, k_d)^2,$$

and $\omega_{eq}^{\otimes d}(x) = \pi^{-d}(1-x^2)^{-1/2}$.

- As you would probably have seen in PO Amblard's talk[4], for $\mu = \mathrm{d}x$, assumptions can be relaxed and K be taken such that $\mathrm{K}(x,x) \propto 1$.

[4] **CoMaAm21**.

Prologue: Determinantal point processes

Numerical integration

Tight interpolation rates in RKHSs

- Consider the RKHS $\mathcal{F}$ with kernel κ, i.e. the completion of

$$\left\{\sum_{i=1}^{M} \alpha_i \kappa(x_i, \cdot), M \in \mathbb{N}, \alpha_1, \ldots, \alpha_n \in \mathbb{R}, x_1, \ldots, x_M \in \mathbb{R}^d\right\}.$$

for the inner product defined by $\langle \kappa(x, \cdot), \kappa(y, \cdot)\rangle_{\mathcal{F}} := \kappa(x, y)$.

- Under general assumptions, $\mathcal{F} \subset L^2(\mathrm{d}\mu)$, is dense, there is an ON basis (e_n) of $L^2(\mathrm{d}\mu)$ and $\sigma_n \to 0$ such that, pointwise,

$$\kappa(x, y) = \sum_{n \geqslant 1} \sigma_n e_n(x) e_n(y).$$

- In that case, $f \in \mathcal{F}$ if and only if $\sum_n \sigma_n^{-1} |\langle f, e_n\rangle|^2$ converges.

- Consider the RKHS $\mathcal{F}$ with kernel κ, i.e. the completion of

$$\left\{ \sum_{i=1}^{M} \alpha_i \kappa(x_i, \cdot), M \in \mathbb{N}, \alpha_1, \dots, \alpha_n \in \mathbb{R}, x_1, \dots, x_M \in \mathbb{R}^d \right\}.$$

for the inner product defined by $\langle \kappa(x, \cdot), \kappa(y, \cdot) \rangle_{\mathcal{F}} := \kappa(x, y)$.

- Under general assumptions, $\mathcal{F} \subset L^2(\mathrm{d}\mu)$, is dense, there is an ON basis (e_n) of $L^2(\mathrm{d}\mu)$ and $\sigma_n \to 0$ such that, pointwise,

$$\kappa(x, y) = \sum_{n \geqslant 1} \sigma_n e_n(x) e_n(y).$$

- In that case, $f \in \mathcal{F}$ if and only if $\sum_n \sigma_n^{-1} |\langle f, e_n \rangle|^2$ converges.

- Consider the RKHS $\mathcal{F}$ with kernel κ, i.e. the completion of

$$\left\{ \sum_{i=1}^{M} \alpha_i \kappa(x_i, \cdot), M \in \mathbb{N}, \alpha_1, \dots, \alpha_n \in \mathbb{R}, x_1, \dots, x_M \in \mathbb{R}^d \right\}.$$

  for the inner product defined by $\langle \kappa(x, \cdot), \kappa(y, \cdot) \rangle_{\mathcal{F}} := \kappa(x, y)$.

- Under general assumptions, $\mathcal{F} \subset L^2(\mathrm{d}\mu)$, is dense, there is an ON basis (e_n) of $L^2(\mathrm{d}\mu)$ and $\sigma_n \to 0$ such that, pointwise,

$$\kappa(x, y) = \sum_{n \geqslant 1} \sigma_n e_n(x) e_n(y).$$

- In that case, $f \in \mathcal{F}$ if and only if $\sum_n \sigma_n^{-1} |\langle f, e_n \rangle|^2$ converges.

- Let $f \in \mathcal{F}$, $g \in L^2(\mathrm{d}\mu)$ then

$$\left| \int fg\mathrm{d}\mu - \sum_{i=1}^{N} w_i f(x_i) \right| \leqslant \|f\|_{\mathcal{F}} \left\| \mu_g - \sum_{i=1}^{N} w_i \kappa(x_i, .) \right\|_{\mathcal{F}}, \qquad (1)$$

  where

$$\mu_g = \int g(x)\kappa(x, .)\mathrm{d}\mu(x)$$

  is the mean element of g.

- Once the nodes $x_1, \ldots, x_N$ are known, minimizing the RHS of (1) in w boils down to inverting an $N \times N$ matrix.

Quadrature and approximation in an RKHS

- Let $f \in \mathcal{F}$, $g \in L^2(\mathrm{d}\mu)$ then

$$\left| \int fg \mathrm{d}\mu - \sum_{i=1}^{N} w_i f(x_i) \right| \leqslant \|f\|_{\mathcal{F}} \left\| \mu_g - \sum_{i=1}^{N} w_i \kappa(x_i, .) \right\|_{\mathcal{F}}, \qquad (1)$$

where

$$\mu_g = \int g(x) \kappa(x, .) \mathrm{d}\mu(x)$$

is the mean element of g.

- Once the nodes $x_1, \dots, x_N$ are known, minimizing the RHS of (1) in w boils down to inverting an $N \times N$ matrix.

Remember $\kappa(x,y) = \sum_{n\geqslant 1} \sigma_n e_n(x) e_n(y)$.

Algorithm 1: DPP

- Take $\mathrm{K}(x,y) = \sum_{n=1}^{N} e_n(x) e_n(y)$.
- Let $x_1, \dots, x_N \sim 1/N! \det[\mathrm{K}(x_i, x_j)] \, \mathrm{d}\mu(x_1) \dots \mathrm{d}\mu(x_N)$.
- Solve the linear problem for the weights $w_1, \dots, w_N$.

Theorem (Belhadji, Bardenet, and Chainais 2019)

Assume $\sum_{n=1}^{N} |\langle g, e_n \rangle|^2 \leqslant 1$. *Let* $r_N = \sum_{m \geqslant N+1} \sigma_m$, *then*

$$\mathbb{E} \left\| \mu_g - \sum_{i=1}^{N} w_i \kappa(x_i, \cdot) \right\|_{\mathcal{F}}^2 \leqslant 2\sigma_{N+1} + 2 \left(N r_N + \sum_{\ell=2}^{N} \frac{\sigma_1}{\ell!^2} \left(\frac{N r_N}{\sigma_1} \right)^{\ell} \right).$$

Remember $\kappa(x,y) = \sum_{n\geqslant 1} \sigma_n e_n(x) e_n(y)$.

Algorithm 1: DPP

- Take $\mathrm{K}(x,y) = \sum_{n=1}^{N} e_n(x)e_n(y)$.
- Let $x_1, \dots, x_N \sim 1/N! \det[\mathrm{K}(x_i,x_j)] \,\mathrm{d}\mu(x_1)\dots\mathrm{d}\mu(x_N)$.
- Solve the linear problem for the weights $w_1, \dots, w_N$.

Theorem (Belhadji 2021)

Assume $\|g\|_\omega \leqslant 1$. Let $r_N = \sum_{m\geqslant N+1} \sigma_m$, then

$$\mathbb{E}\left\|\mu_g - \sum_{i=1}^{N} w_i \kappa(x_i,\cdot)\right\|_{\mathcal{F}}^2 \leqslant 4 r_N.$$

2nd attempt: volume sampling and tight rates

Algorithm 2: volume sampling

- Let $x_1, \dots, x_N \sim Z^{-1} \det[\kappa(x_i, x_j)] \,\mathrm{d}\mu(x_1) \dots \mathrm{d}\mu(x_N)$
- Again, solve the linear program for the weights $w_1, \dots, w_N$.

Theorem (Belhadji, Bardenet, and Chainais 2020b)

Assume again $\sum_{n=1}^{N} |\langle g, e_n \rangle|^2 \leqslant 1$. *Then*

$$\mathbb{E}\left\| \mu_g - \sum_{i=1}^{N} w_i \kappa(x_i, \cdot) \right\|_{\mathcal{F}}^2 \leqslant \sigma_N (1 + \beta_N),$$

where $\beta_N = \min_{M \in [2:N]} [(N - M + 1)\sigma_N]^{-1} \sum_{m \geqslant M} \sigma_m$.

- It is known[5] that $\inf_{\substack{Y \subset \mathcal{F} \\ \dim Y = N}} \sup_{\|g\|_\omega \leqslant 1} \inf_{y \in Y} \|\mu_g - y\|_{\mathcal{F}}^2 = \sigma_{N+1}$.

[5] Pinkus 2012.

2nd attempt: volume sampling and tight rates

Algorithm 2: volume sampling

- Let $x_1, \ldots, x_N \sim Z^{-1} \det[\kappa(x_i, x_j)] \, \mathrm{d}\mu(x_1) \ldots \mathrm{d}\mu(x_N)$
- Again, solve the linear program for the weights $w_1, \ldots, w_N$.

Theorem (Belhadji, Bardenet, and Chainais 2020b)

Assume again $\sum_{n=1}^{N} |\langle g, e_n \rangle|^2 \leqslant 1$. *Then*

$$\mathbb{E}\left\| \mu_g - \sum_{i=1}^{N} w_i \kappa(x_i, \cdot) \right\|_{\mathcal{F}}^2 \leqslant \sigma_N \left(1 + \beta_N\right),$$

where $\beta_N = \min_{M \in [2:N]} [(N - M + 1)\sigma_N]^{-1} \sum_{m \geqslant M} \sigma_m$.

- It is known[5] that $\inf_{\substack{Y \subset \mathcal{F} \\ \dim Y = N}} \sup_{\|g\|_\omega \leqslant 1} \inf_{y \in Y} \|\mu_g - y\|_{\mathcal{F}}^2 = \sigma_{N+1}$.

[5] Pinkus 2012.

Volume sampling is a mixture of projection DPPs

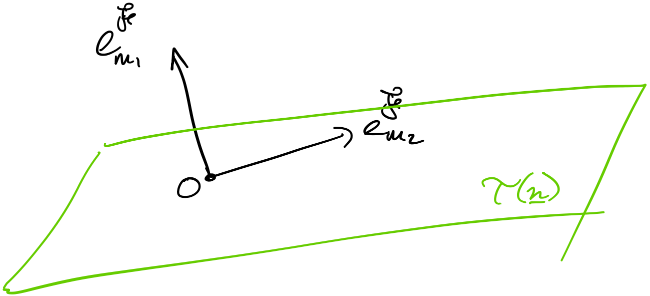
- Remember $\kappa(x, y) = \sum_{n \geqslant 1} \sigma_n e_n(x) e_n(y)$.
- For $U \subset \mathbb{N}^*$ define the projection kernel

$$K_U(x, y) = \sum_{u \in U} e_u(x) e_u(y). \tag{2}$$

For $N \in \mathbb{N}^*$, we have

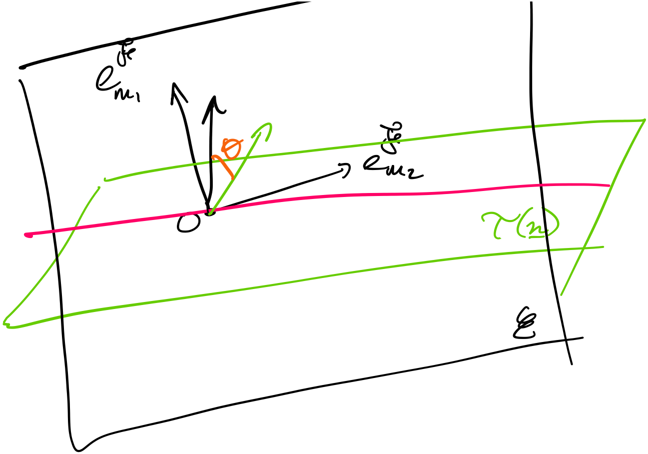
$$\det \kappa(x_i, x_j) \propto \sum_{|U|=N} \left(\prod_{u \in U} \sigma_u \right) \frac{1}{N!} \det(K_U(x_i, x_j)). \tag{3}$$

- Remember $\kappa(x,y) = \sum_{n\geqslant 1} \sigma_n e_n(x) e_n(y)$.
- Let $(e_n^{\mathcal{F}} = \sqrt{\sigma_n} e_n)$ be ON in $\mathcal{F}$.



Some graphical intuition on volume sampling

- Remember $\kappa(x,y) = \sum_{n\geqslant 1} \sigma_n e_n(x) e_n(y)$.
- Let $(e_n^{\mathcal{F}} = \sqrt{\sigma_n} e_n)$ be ON in $\mathcal{F}$.



Many open problems

- Robustness to RKHS hypothesis / model choice.
- Practical relevance of RKHS hypothesis.
- What should $g \in L^2(\mu)$ be in $\int fg \mathrm{d}\mu$?
- How do we efficiently sample from continuous volume sampling without spectral knowledge? See e.g. Rezaei and Gharan 2019.
- Kernel interpolation is similar to column-subset selection for linear regression[6], where DPPs and VS yield similar bounds[7].

[6] Derezinski and Mahoney 2020.
[7] Belhadji, Bardenet, and Chainais 2020a.

References I

Bardenet, R. and A. Hardy (2020). "Monte Carlo with Determinantal Point Processes". In: *Annals of Applied Probability.*

Belhadji, A. (2021). "An analysis of Ermakov-Zolotukhin quadrature using kernels". In: *Advances in Neural Information Processing Systems (NeurIPS).*

Belhadji, A., R. Bardenet, and P. Chainais (2019). "Kernel quadrature with determinantal point processes". In: *Advances in Neural Information Processing Systems (NeurIPS).*

– (2020a). "A determinantal point process for column subset selection". In: *Journal of Machine Learning Research (JMLR).*

– (2020b). "Kernel interpolation with continuous volume sampling". In: *International Conference on Machine Learning (ICML).*

Derezinski, M. and M. Mahoney (2020). "Determinantal Point Processes in Randomized Numerical Linear Algebra". In: *arXiv preprint arXiv:2005.03185.*

Gautier, G., R. Bardenet, G. Polito, and M. Valko (2019). "DPPy: Sampling Determinantal Point Processes with Python". In: *Journal of Machine Learning Research; Open Source Software (JMLR MLOSS).*

Hough, J. B., M. Krishnapur, Y. Peres, and B. Virág (2006). "Determinantal processes and independence". In: *Probability surveys.*

Lavancier, F., J. Møller, and E. Rubak (2014). "Determinantal point process models and statistical inference". In: *Journal of the Royal Statistical Society, Series B.* B.

References II

Macchi, O. (1975). "The coincidence approach to stochastic point processes". In: *Advances in Applied Probability* 7, pp. 83–122.

Pinkus, A. (2012). *N-widths in Approximation Theory*. Vol. 7. Springer Science & Business Media.

Rezaei, A. and S. O. Gharan (2019). "A Polynomial Time MCMC Method for Sampling from Continuous Determinantal Point Processes". In: *International Conference on Machine Learning*, pp. 5438–5447.

Soshnikov, A. (2002). "Gaussian Limit for Determinantal Random Point Fields". In: *Annals of Probability* 30.1, pp. 171–187.